Multi-ancestry genome-wide association study identifies new asthma susceptibility loci that co-localize with immune cell enhancer histone marks

Florence Demenais[1,2*], Patricia Margaritte-Jeannin[1,2*], Kathleen C Barnes[3*], William OC Cookson[4*], Janine Altmueller[5], Wei Ang[6], R Graham Barr[7], Allan B Becker[8], John Beilby[9], Hans Bisgaard[10], Unnur Steina Bjornsdottir[11], Eugene Bleecker[12], Klaus Bønnelykke[10], Dorret I Boomsma[13], Emmanuelle Bouzigon[1,2], Christopher E Brightling[14], Myriam Brossard[1,2], Guy G Brusselle[15,16,17], Esteban Burchard[18], Kristin M Burkart[19], Andrew Bush[20,21], Moira Chan-Yeung[22], Kian Fan Chung[20,23], Alexessander Couto Alves[24], John A Curtin[25], Adnan Custovic[26], Denise Daley[22], Johan C de Jongste[27], Blanca del Rio Navarro[28], Kathleen M Donohue[7], Liesbeth Duijts[29], Celeste Eng[30], Johan G Eriksson[31], Martin Farrall[32,33], Yuliya Fedorova[34], Bjarke Feenstra[35], Manuel A Ferreira[36], Australian Asthma Genetics Consortium collaborators, Maxim B Freidin[37], Zofia Gajdos[38,39], Jim Gauderman[40], Ulrike Gehring[41], Frank Geller[35], Jon Genuneit[42], Sina A Gharib[43], Frank Gilliland[40], Raquel Granell[44,45], Penelope Graves[46], Daniel F Gudbjartsson[47,48], Tari Haahtela[49], Susan R Heckbert[50], Dick Heederik[51], Joachim Heinrich[52,53], Markku Heliövaara[54], John Henderson[44,45], Blanca E Himes[55], Hiroshi Hirose[56], Joel N Hirschhorn[57,58,59], Albert Hofman[16], Patrick Holt[60], Jouke Jan Hottenga[13], Thomas J Hudson[61], Jennie Hui[62], Medea Imboden[63,64], Vladimir Ivanov[65], Vincent WV Jaddoe[66], Alan James[67,68], Christer Janson[69], Marjo-Riitta Jarvelin[70,71,72,73], Deborah Jarvis[20,74], Graham Jones[75], Ingileif Jonsdottir[47,76], Bonnie R Joubert[77], Pekka Jousilahti[54], Michael Kabesch[78], Mika Kähönen[79], David B Kantor[80,81], Alexandra S Karunas[34,82], Elza Khusnutdinova[34,82], Gerard H Koppelman[83,84], Anita L Kozyrskyj[85], Eskil Kreiner[10], Michiaki Kubo[86], Rajesh Kumar[87,88], Ashish Kumar[63,64,89], Mikko Kuokkanen[54,90], Lies Lahousse[15,16], Tarja Laitinen[91], Catherine Laprise[92], Mark Lathrop[93], Susanne Lau[94], Young-Ae Lee[95,96], Terho Lehtimäki[97], Sébastien Letort[1,2], Albert M Levin[98], Guo Li[99], Liming Liang[100], Laura R Loehr[101], Stephanie J London[102], Daan W Loth[16], Ani Manichaikul[103], Ingo Marenholz[95,96], Fernando Martinez[46], Melanie C Matheson[104], Kenji Matsumoto[105], Hamdi Mbarek[13], Wendy L McArdle[106], Mads Melbye[35], Erik Melén[89,107,108], Deborah Meyers[12], Sven Michel[78], Hamida Mohamdi[1,2], Arthur W Musk[109,110], Rachel Myers[111], Maartje A E Nieuwenhuis[84,112], Emiko Noguchi[113], George T O'Connor[114,115], Ludmila M Ogorodova[116], Cameron D Palmer[57,59], Aarno Palotie[90,117], Julie E Park[118], Craig E Pennell[6], Göran Pershagen[89,107], Alexey Polonikov[65], Dirkje S Postma[84,112], Nicole Probst-Hensch[63,64], Valery P Puzyrev[37], Benjamin A Raby[119], Olli T Raitakari[120], Adaikalavan Ramasamy[24,121], Stephen S Rich[103], Colin F Robertson[122], Isabelle Romieu[123], Muhammad T Salam[40], Veikko Salomaa[54], Vivi Schlünssen[124], Robert Scott[125], Polina A Selivanova[126],

Torben Sigsgaard[124], Angela Simpson[25], Valérie Siroux[127,128], Lewis J Smith[129], Maria Solodilova[65], Marie Standl[53], Kari Stefansson[47,76], David P Strachan[130], Bruno H Stricker[16,131,132], Atsushi Takahashi[86], Philip J Thompson[133], Gudmar Thorleifsson[47], Unnur Thorsteinsdottir[47,76], Carla MT Tiesler[53,134], Dara G Torgerson[30], Tatsuhiko Tsunoda[86,135], André G Uitterlinden[132], Ralf JP van der Valk[136], Amaury Vaysse[1,2], Sailaja Vedantam[38,39], Andrea Von Berg[137], Erika Von Mutius[138,139], Judith M Vonk[84,140], Johannes Waage[10], Nick J Wareham[125], Scott T Weiss[119], Wendy B White[141], Magnus Wickman[89,108], Elisabeth Widén[90], Gonneke Willemsen[13], L Keoki Williams[142,143], Inge M Wouters[51], James J Yang[144], Jing Hua Zhao[125], Miriam F Moffatt[4*], Carole Ober[145*] & Dan L Nicolae[146*]

*These authors contributed equally to this work

[1]Genetic Variation and Human Diseases Unit, UMR-946, INSERM, Paris, France.
[2]Institut Universitaire d'Hématologie, Université Paris Diderot, Université Sorbonne Paris Cité, Paris, France.
[3]Division of Biomedical Informatics and Personalized Medicine, Colorado Center for Personalized Medicine, University of Colorado, Denver, Colorado, USA.
[4]Section of Genomic Medicine, National Heart and Lung Institute, London, UK.
[5]Cologne Center for Genomics (CCG), University of Cologne, Cologne, Germany.
[6]School of Women's and Infants' Health, The University of Western Australia, Perth, Western Australia, Australia.
[7]Departments of Medicine and Epidemiology, Columbia University, New York, New York, USA.
[8]Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Manitoba, Canada.
[9]Department of the Diagnostic Genomics Laboratory, PathWest Laboratory Medicine, Queen Elizabeth II Medical Centre, Nedlands, Western Australia, Australia.
[10]COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark.
[11]Department of Medicine, Landspitali, The National University Hospital of Iceland, Reykjavik, Iceland.
[12]Center for Genomics, Wake Forest University School of Medicine, Winston Salem, North Carolina, USA.
[13]Department of Biological Psychology, Vrjie Universiteit Amsterdam and EMGO Institute for Health and Care Research, Amsterdam, The Netherlands.
[14]University of Leicester, Institute for Lung Health, Glenfield Hospital, Leicester, UK.
[15]Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium.
[16]Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands.
[17]Department of Respiratory Medicine, Erasmus MC, Rotterdam, The Netherlands.

[18]Department of Bioengineering & Therapeutic Sciences and Medicine, University of California, San Francisco, California, USA.
[19]Division of Pulmonary, Allergy and Critical Care, College of Physicians and Surgeons, Columbia University, New York, New York, USA.
[20]National Heart and Lung Institute, Imperial College London, London, UK.
[21]Royal Brompton Harefield NHS Trust, London, UK.
[22]Department of Medicine, The University of British Columbia, Vancouver, British Columbia, Canada.
[23]Biomedical Research Unit, Royal Brompton & Harefield NHS Trust, London, UK.
[24]Department of Epidemiology and Biostatistics, Imperial College London, London, UK.
[25]Centre for Respiratory Medicine and Allergy, Institute of Inflammation and Repair, Manchester Academic Health Science Centre, The University of Manchester and University Hospital of South Manchester NHS Foundation Trust, Manchester, UK.
[26]Department of Paediatrics, Imperial College London, London, UK.
[27]Department of Pediatrics - Division of Respiratory Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands.
[28]Hospital Infantil Federico Gomez, Mexico City, Mexico.
[29]Department of Pediatrics-Division of Respiratory Medicine and Department of Pediatrics-Division of Neonatology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands.
[30]Department of Medicine, University of California, San Francisco, California, USA.
[31]Department of General Practice and Primary Health Care, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.
[32]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK.
[33]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.
[34]Institute of Biochemistry and Genetics, Ufa Scientific Center of Russian Academy of Sciences, Ufa, Russian Federation.
[35]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark.
[36]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia.
[37]Population Genetics Laboratory, Research Institute of Medical Genetics, Tomsk NRMC, Tomsk, Russian Federation.
[38]Divisions of Genetics and Endocrinology, Children's Hospital, Boston, Massachusetts, USA.
[39]Broad Institute, Cambridge, Massachusetts, USA.
[40]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.

[41]Institute for Risk Assessment Sciences, Division Environmental Epidemiology, Utrecht University, Utrecht, the Netherlands.
[42]Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany.
[43]Department of Medicine, University of Washington, Seattle, Washington, USA.
[44]School of Social and Community Medicine, Faculty of Health Sciences, University of Bristol, Bristol, UK.
[45]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.
[46]Arizona Respiratory Center and BIO5 Institute, The University of Arizona, Tucson, Arizona, USA.
[47]deCODE genetics, Amgen Inc., Reykjavik, Iceland.
[48]School of Engineering and Natural Sciences, Univeristy of Iceland, Reykjavik, Iceland.
[49]Skin and Allergy Hospital, University of Helsinki, Helsinki, Finland.
[50]Department of Epidemiology, University of Washington, Seattle, Washington, USA.
[51]Division Environmental Epidemiology, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands.
[52]Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, University Hospital in Munich, Munich, Germany.
[53]Institute of Epidemiology I, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany.
[54]Department of Health, National Institute for Health and Welfare (THL), Helsinki, Finland.
[55]Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, Pennsylvania, USA.
[56]Health Center, Department of Internal Medicine, Keio University, Tokyo, Japan.
[57]Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, Massachusetts, USA.
[58]Departments of Pediatrics and Genetics, Harvard Medical School, Boston, Massachusetts, USA.
[59]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.
[60]Cell Biology Telethon Kids Institute, The University of Western Australia, Subiaco, Western Australia, Australia.
[61]Ontario Institute for Cancer Research, Toronto, Ontario, Canada.
[62]Busselton Population Medical Research Institute, Perth, Western Australia, Australia.
[63]Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland.
[64]University of Basel, Basel, Switzerland.
[65]Department of Biology, Medical Genetics and Ecology, Kursk State Medical University, Kursk, Russian Federation.
[66]The Generation R Study Group, Department of Pediatrics and Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands.

[67]Department of Pulmonary Physiology and Sleep Medicine, Busselton Population Medical Research Institute, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia.
[68]School of Population Health, University of Western Australia, Crawley, Western Australia, Australia.
[69]Department of Medical Sciences, University of Uppsala, Uppsala, Sweden.
[70]Department of Epidemiology and Biostatistics, MRC–PHE Centre for Environment & Health, School of Public Health, Imperial College London, London, UK.
[71]Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland.
[72]Biocenter Oulu, University of Oulu, Oulu, Finland.
[73]Unit of Primary Care, Oulu University Hospital, Oulu, Finland.
[74]MRC-PHE Centre for Environment and Health, Imperial College, London, UK.
[75]School of Science and Health, Western Sydney University, Sydney, New South Wales, Australia.
[76]Faculty of Medicine, Univeristy of Iceland, Reykjavik, Iceland.
[77]National Institute of Envirionmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, USA.
[78]Department of Pediatric Pneumology and Allergy, University Children's Hospital Regensburg (KUNO), Regensburg, Germany.
[79]Department of Clinical Physiology, University of Tampere and Tampere University Hospital, Tampere, Finland.
[80]Department of Anesthesiology, Perioperative and Pain Medicine, Division of Critical Care Medicine, Boston Children's Hospital, Boston, Massachusetts, USA.
[81]Department of Anaesthesia, Harvard Medical School, Boston, Massachusetts, USA.
[82]Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa, Russian Federation.
[83]University of Groningen, University Medical Center Groningen, Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, Groningen, the Netherlands.
[84]GRIAC Research Institute, Groningen, the Netherlands.
[85]Department of Pediatrics, The University of Alberta, Alberta, Alberta, Canada.
[86]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.
[87]Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois, USA.
[88]Divison of Allergy and Clinical Immunology, Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA.
[89]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.
[90]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.
[91]Department of Pulmonary Medicine, Turku University Hospital, University of Turku and Turku University, Turku, Finland.

[92]Département des Sciences Fondamentales, Université du Québec à Chicoutimi, Chicoutimi, Quebec, Canada.
[93]McGill University and Genome Quebec Innovation Centre, Montréal, Quebec, Canada.
[94]Pediatric Pneumology and Immunology, Charité, Berlin, Germany.
[95]Max-Delbrück-Centrum (MDC) for Molecular Medicine, Berlin, Germany.
[96]Pediatric Allergology, Experimental and Clinical Research Center, Charité, Berlin, Germany.
[97]Department of Clinical Chemistry, Fimlab Laboratories, University of Tampere Faculty of Medicine and Life Sciences, Tampere, Finland.
[98]Department of Public Health Sciences, Henry Ford Health System, Detroit, Michigan, USA.
[99]Department of Medicine, University of Washington, Seattle, Washington, USA.
[100]Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.
[101]Division of General Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.
[102]National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, USA.
[103]Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA.
[104]Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia.
[105]Department of Allergy and Clinical Immunology, National Research Institute for Child Health and Development, Tokyo, Japan.
[106]Bristol Bioresource Laboratories, School of Social and Community Medicine, University of Bristol, Bristol, UK.
[107]Centre for Occupational and Environmental Medicine, Stockholm County Council, Stockholm, Sweden.
[108]Sachs Children's Hospital, Stockholm, Sweden.
[109]Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia.
[110]schools of Population Health and Medicine and Pharmacology, University of Western Australia, Perth, Western Australia, Australia.
[111]Department of Medicine, Duke University, Durham, North Carolina, USA.
[112]University of Groningen, University Medical Center Groningen, Department of Pulmonology, Groningen, the Netherlands.
[113]Department of Medical Genetics, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan.
[114]Pulmonary Center, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA.

[115]The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA.
[116]Department of Faculty Pediatrics, Siberian State Medical University, Tomsk, Russian Federation.
[117]The Wellcome Trust Sanger Insitute, Cambridge, UK.
[118]The University of British Columbia, Vancouver, British Columbia, Canada.
[119]Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA.
[120]Department of Clinical Physiology and Nuclear Medicine, University of Turku and Turku University Hospital, Turku, Finland.
[121]Department of Medical and Molecular Genetics, King's College London, London, UK.
[122]Respiratory Medicine, Murdoch Children's Research Institute, Melbourne, Victoria, Australia.
[123]International Agency for Research on Cancer, Lyon, France.
[124]Department of Public Health, Section for Environment, Occupation & Health, Aarhus University, Aarhus, Denmark.
[125]MRC Epidemiology Unit & Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK.
[126]Department of Faculty Therapy, Siberian State Medical University, Tomsk, Russian Federation.
[127]Inserm U1209, Institute for Advanced Biosciences, Team of Environmental Epidemiology applied to Reproduction and Respiratory Health, Grenoble, France.
[128]Université de Grenoble Alpes / CNRS UMR5309, Institute for Advanced Biosciences, Team of Environmental Epidemiology applied to Reproduction and Respiratory Health, Grenoble, France.
[129]Division of Pulmonary and Critical Care Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA.
[130]Population Health Research Institute, St George's University of London, London, UK.
[131]Netherlands Healthcare Inspectorate, The Hague, The Netherlands.
[132]Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands.
[133]Institute for Respiratory Health, Harry Perkins Institute of Medical Research and The Lung Health Clinic, Nedlands, Western Australia, Australia.
[134]Division of Metabolic Diseases and Nutritional Medicine, Dr. von Hauner Children's Hospital, Ludwig-Maximilians-University of Munich, Munich, Germany.
[135]Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan.
[136]The Generation R Study Group, Department of Pediatrics - Division of Respiratory Medicine and Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands.
[137]Department of Pediatrics, Marien-Hospital Wesel, Wesel, Germany.
[138]Dr. Von Hauner Children's Hospital, Ludwig Maximilians University Munich, Munich, Germany.

[139]Member of the German Center for Lung Research, Germany.
[140]University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, the Netherlands.
[141]Undergraduate Training and Education Center (UTEC), Jackson Heart Study, Tougaloo College, Jackson, Mississippi, USA.
[142]Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit, Michigan, USA.
[143]Department of Internal Medicine, Henry Ford Health System, Detroit, Michigan, USA.
[144]School of Nursing, University of Michigan, Ann Arbor, Michigan, USA.
[145]Department of Human Genetics, The University of Chicago, Chicago, Illinois, USA.
[146]Departments of Statistics, Human Genetics and Medicine, Section of Genetic Medicine, The University of Chicago, Chicago, Illinois, USA.

Correspondence should be sent to:

Florence Demenais
Genetic Variation and Human Diseases Unit
INSERM, UMR-946
27 rue Juliette Dodu
75010 Paris, France
Phone : +33 1 53 72 50 02 – Fax : +33 1 53 72 50 49
e-mail: florence.demenais@inserm.fr

or

Dan Nicolae
Departments of Statistics
The University of Chicago
5734 S. University Avenue
Chicago, IL 60637, USA
Phone: +1 773 702 4837 - Fax: +1 773 702 9810
e-mail: nicolae@galton.uchicago.edu

Abstract

We examined common variation in asthma risk by conducting the largest meta-analysis to date of asthma genome-wide association studies (23,948 cases, 118,538 controls) from multi-ancestry populations. We identified five new asthma loci, uncovered two novel associations at two known asthma loci, established asthma associations at two loci implicated previously in comorbidity of asthma plus hay fever, and confirmed nine known loci. Investigation of pleiotropy showed large overlaps in genetic variants with autoimmune and inflammatory diseases. Enrichment of asthma loci near enhancer marks, especially in immune cells, suggests a major role of these loci in the regulation of immune-related mechanisms.

Asthma is a complex disease that affects hundreds of millions of people worldwide. The prevalence of asthma varies between populations and ethnicities, ranging in the U.S. from 3.9% in Mexican Americans to 12.5% in African Americans[1]. The contribution of genetic factors to asthma risk has been demonstrated in family studies, where heritability estimates range from 25% to 80%[2]. The large variability in prevalence and heritability estimates reflects the significant role of environmental exposures on disease risk and the considerable phenotypic heterogeneity that is a hallmark of this disease. These features largely explain why genome-wide association studies (GWAS) of asthma have uncovered a smaller number of loci than similarly sized studies of other multifactorial diseases[3]. Indeed, up to now, only 21 loci have been associated with asthma *per se* in 20 studies, and these loci explain only part of the genetic risk. Moreover, an exome-chip study of EVE consortium samples showed no evidence for large-effect of low frequency or rare variants in asthma risk[4]. This lack of rare variants in asthma is similar to findings from other common diseases[5,6]. To generate larger sample sizes for GWAS meta-analysis of asthma enabling the discovery of novel risk loci, we established the Trans-National Asthma Genetic Consortium (TAGC) across worldwide groups of investigators with genome-wide data. The TAGC sample of more than 142,000 individuals representing diverse ancestry allowed us to construct a comprehensive catalog of asthma risk variants that are robust across populations and environmental exposures. By combining TAGC meta-analysis results with existing databases, we have been able to assess the genetic architecture of asthma risk alleles with respect to functional effects and shared effects with other diseases.

RESULTS

Meta-analysis of asthma GWAS from ethnically diverse populations

We combined data from asthma GWAS with high-density genotyped and imputed SNP data (2.93 million SNPs) in the following populations: European-ancestry (19,954 cases, 107,715

controls), African-ancestry (2,149 cases, 6,055 controls), Japanese (1,239 cases, 3,976 controls) and Latino (606 cases, 792 controls) (Supplementary Table 1). After extensive QC of summary data provided by each participating group (Online Methods, Supplementary Note and Supplementary Table 2), we first conducted ancestry-specific meta-analyses followed by a multi-ancestry meta-analysis of all populations (23,948 cases, 118,538 controls) to identify additional loci with pan-ancestry effects. As childhood-onset asthma may be distinct from later-onset asthma[7] and may represent a more homogeneous subgroup, we also performed analyses on the pediatric subgroup (asthma onset ≤16 years; 8,976 cases, 18,399 controls). All meta-analyses were performed using both fixed-effects (significance summarized in P_{fixed}) and random-effects (P_{random}). For all results, we employed a stringent P_{random} (or P_{fixed}) threshold of 10^{-8} to define genome-wide significance. As results were consistent between methods regarding the detection of loci with at least one SNP significantly associated with asthma, we present the results from the random-effects analysis for the European-ancestry and multi-ancestry meta-analyses which include a large number of studies allowing an accurate estimate of the between-study variance and results from the fixed-effects analysis for the African-ancestry, Japanese and Latino meta-analyses (including at most seven studies). We observed little evidence of inflation in the test statistics in either the ancestry-specific (lambda: European-ancestry, 1.031; African-ancestry, 1.014; Japanese, 1.021; Latino, 1.044) or multi-ancestry (lambda=1.046) meta-analyses (Supplementary Fig.1).

We identified 523 genome-wide significant SNPs ($P_{random} \leq 10^{-8}$) at 14 loci in European-ancestry populations (Fig.1, Table 1, Supplementary Tables 3 and 4; see definition of a locus in Online Methods). No genome-wide significant risk loci were detected in ancestry-specific analyses within the African-ancestry, Japanese or Latino populations (Supplementary Fig.2 and Supplementary Tables 5-7), possibly due to a lack of power. In the combined multi-ancestry meta-analysis, 90 additional SNPs of which 20 belonged to four additional regions were

significant at $P_{random} \leq 10^{-8}$ (Fig.2, Table 1, and Supplementary Tables 3 and 8). Altogether, the 18 chromosomal regions included: five new loci associated with asthma at 5q31.3, 6p22.1, 6q15, 12q13.3, 17q21.33, two new significant associations at 6p21.33 and 10p14 that were independent from previously reported signals at these loci in ancestry-specific populations (Latino[8] and Japanese[9]), two reported associations with the phenotype asthma plus hay fever at 8q21.13 and 16p13.13 in one European-ancestry study[10], and nine previously identified asthma loci. The two asthma plus hay fever signals on 8q21.13 and 16p13.13 were not detected with asthma alone in the previous study[10]; in the current study, the 8q21.13 locus reached genome-wide significance only in the multi-ancestry analysis. None of the lead SNPs at these 18 loci showed evidence for heterogeneity in effect sizes across studies except for the lead variant at 9p24.1 (P_{het} for Cochran's Q test[11]=0.008 across European-ancestry studies and P_{het}=0.02 across multi-ancestry studies, Table 1, Supplementary Fig.3). Heterogeneity in the ancestry-specific effect sizes of the lead SNPs was only detected at the 6p22.1 and 6p21.33 loci (P_{ethnic} for heterogeneity between the four ancestry-specific odds-ratios equal to 0.003 for rs1233578 at 6p22.1 and to 0.02 for rs3131064 at 6p21.33), where the lead SNPs were only significant in European-ancestry populations (Supplementary Table 3). The meta-analysis of the pediatric subgroup showed strong evidence for association at five of the 18 loci (2q12, 5q31 and 17q12-21 having $P_{random} \leq 10^{-8}$; 6p21.33 and 9p24.1 having $P_{random} \leq 5x10^{-8}$) (Supplementary Figs.4 and 5 and Supplementary Table 9). No novel loci specific to that group were identified.

Besides the detection of new signals, our GWAS provided genome-wide significant confirmation of nine previously reported loci both in European-ancestry and multi-ancestry meta-analyses (Table 1; Supplementary Fig.6). It is of note that the lead SNP (rs2952156) at the 17q12-21 locus is within *ERBB2* ($P_{random}=2.2x10^{-30}$ in multi-ancestry meta-analysis), which is at least 180kb from published asthma-associated signals within the *GSDMB/ORMDL3* haplotype block[3] (Supplementary Fig.7). This is due to effect size heterogeneity across studies

(0.001≤P_{het}≤0.05) that extends over a 200kb region spanning *ORMDL3* and *GSDMB* (Supplementary Table 10). This heterogeneity is partly due to age of onset of asthma, as previously reported[7]. Indeed, in the pediatric group, the 17q12-21 SNPs did not show heterogeneity (P_{het}≥0.09) and the lead SNP rs8069176 is 3.6 kb proximal to G*SDMB* (P_{random}=P_{fixed}=4.4 x10^{-26}). This is supported by the comparison in SNP effect size between pediatric and non-pediatric studies which shows a significant difference for rs8069176 (P_{het}=7.4x10^{-4}) but no difference for rs2952156 (P_{het}=0.11). The two SNPs, rs2952156 and rs8069176, are in moderate LD (r^2=0.30). Interestingly, both SNPs are in strong LD (r^2>0.9) with missense variants that lie in *ERBB2* (for the proxy of rs2952156) and in *ZPBP2* and *GSDMB* (for the proxies of rs8069176). In addition, rs2952156 and rs8069176 are associated with *GSDMB* and *ORMDL3* expression in five blood e-QTL databases[12-15] and with expression of *GSDMA*, *CDK12*, *GSDMB*, *ORMDL3* in lung[14,16]. However, only rs2952156 is associated with *PGAP3* transcript in lung[14,16] (Supplementary Table 11a). Further exploration of eQTL data from GTEx[14], which includes both blood and lung samples, indicated that rs8069176 accounted for a large part of the association of the most significant SNP with *ORMDL3* transcript in blood while rs2952156 accounted for a large part of the association of the most significant SNP with *PGAP3* transcript in lung (Supplementary Table 11b). This suggests that the asthma-associated signals near *PGAP3*/*ERBB2* and *ORMDL3*/*GSDMB* blocks may affect asthma risk through the expression of different genes in different tissues, as also suggested in a study of lung epithelial cells[17]. However, these results need to be confirmed in additional samples and integration of genotypic, gene expression and epigenetic data may bring further insight into the mechanisms underlying asthma risk at this locus.

Finally, out of the 21 published asthma loci, 12 loci were not confirmed by this study (Supplementary Table 12). The strongest asthma-associated SNP at each of these non-replicated loci, as reported in the GWAS catalog[3], had *P*-values ranging between 0.99 and 8.5x10^{-7} in our

meta-analyses. The non-replication of these loci is due to the following reasons: (i) three loci were found in a study with small sample size and the SNPs had very low allele frequency[18]; (ii) four loci were ancestry-specific[4,9,19]; (iii) three loci were detected in studies with a small number of cases (<1,200 subjects) at the discovery stage[20-22]. For the other two loci reported in European-ancestry studies[23,24], *IL6R* rs4129267 and *IL2RB* rs2284033 ($P_{random=}$ $8.5x10^{-7}$ and $7.9\ x10^{-5}$ in multi-ancestry meta-analysis) had their *P*-values dropping to 0.01 and 0.08 respectively after removing the datasets of the initial publications[23,24].

To investigate whether the 18 asthma loci identified in this study contain multiple distinct signals, we performed approximate conditional regression analysis based on summary statistics for all loci, except for the 9p24.1 region which showed heterogeneity in SNP effect size across studies over the whole locus. For the 17q12-21 locus, this analysis was restricted to the pediatric sub-group in which there was no heterogeneity. This analysis was based on summary statistics of European-ancestry fixed-effects meta-analysis, using the Genome-wide Complex Trait Analysis (GCTA) software[25] (Online Methods). After conditioning on the lead SNP in each investigated region, four secondary signals (2q12, 5q22.1, 5q31, 6p21.32) remained genome-wide significant ($P_{fixed} \leq 10^{-8}$) after conditioning on the lead SNP (Supplementary Table 13), thus resulting in 22 distinct genome-wide significant signals identified by this study.

To provide biological insight into our findings, we conducted a comprehensive bioinformatic assessment of the asthma association signals using several resources. First, to pinpoint the most likely candidate genes at the nine loci harboring novel associations with asthma *per se*, we interrogated results of six eQTL studies in tissues relevant to asthma, blood (including peripheral blood[13,14], lymphoblastoid cell lines (LCLs)[12,15], monocytes[26]) and whole lung tissue[14,16]. This e-QTL investigation was complemented by the search for missense variants potentially tagged by the association signals. To assess the level of overlap of asthma

associations with susceptibility loci for other phenotypes, we next interrogated the GWAS catalog[3] while varying the strength of association with asthma (thresholds from 10^{-8} to 10^{-3}). To get better insight on how asthma associated variants may functionally influence disease, we then interrogated the ROADMAP/ENCODE functional genomics data generated in a wide range of human cell types[27]. Finally, the degree of connectivity between the asthma-associated loci was assessed through text mining[28]. Results from these analyses are described in the following sections.

Potential candidate genes at the nine loci harboring novel associations with asthma

A summary of eQTL analysis for these nine loci is described in Table 2 and Supplementary Table 14.

New asthma susceptibility loci

Five new loci were identified in the current study. The strongest new signal in both European-ancestry (P_{random}=8.6x10^{-13}) and multi-ancestry (P_{random}=2.2x10^{-12}) meta-analyses was with SNP rs2325291 in the intron of *BACH2* at 6q15 (Fig.1d), which is strongly correlated with rs10455168 (r^2=0.91), a cis-eQTL altering expression of *BACH2* in blood[13]. *BACH2* encodes a zip transcription factor that is a key regulator of nucleic acid-triggered antiviral responses in human cells[29]. The second strongest signal was with rs17637472 (P_{random}=3.3x10^{-9} and 6.6x10^{-9} in European-ancestry and multi-ancestry analyses), which lies between *ZNF652* and *PHB* at 17q21.33 (Fig.1f), and is a strong cis-eQTL for *GNGT2* (173kb from rs17637472) in blood[12,13,15,26]. GNGT2 was shown to interact with beta-Arrestin 1 to promote G-protein-dependent Akt signaling for NF-kappaB activation[30]. Among the other new signals, the lead SNP rs1233578 at 6p22.1 (P_{random}=5.3x10^{-9} in European-ancestry populations) resides between *TRIM27* and *GPX5* (Fig.1b). No cis-eQTL was identified at this locus. Notably, TRIM27 was reported to be a negative regulator of CD4 T cells[31] and to negatively regulate signaling

445 mediated by NOD2, a protein involved in antimicrobial responses[32] and a key innate immune
446 receptor in inflammation. The 12q13.3 lead SNP (rs167769), which was only significant in the
447 multi-ancestry analysis (P_{random}=3.9x10^{-9}), lies in an intron of *STAT6* (Fig.2e) and is strongly
448 associated with *STAT6* expression in blood[12,13,15] and lung[16]. STAT6 is a transcription factor
449 that is essential for the functional responses of Th2 lymphocytes mediated by IL-4 and IL-13[33].
450 This result robustly establishes the association of *STAT6* with asthma risk that has been disputed
451 by candidate gene studies[34]. The 5q31.3 lead SNP rs7705042 (P_{random}=7.9x10^{-9} in multi-
452 ancestry analysis) is within an intron of *NDFIP1* (Fig.2b) and associated with *NDFIP1*
453 expression in blood[13,15]. NDFIP1 is a potent inhibitor of antiviral response[35] and inflammation
454 processes[36].

455 ***Novel signals at loci previously associated with asthma in ancestry-specific populations***

456 Two associations in our study were with novel SNPs at loci previously reported to be associated
457 with asthma in Latinos[8] and Japanese[9]. The first one, located at 6p21.33, was reported in an
458 admixture mapping study in Latinos[8]. The lead SNP in our study, rs3131064, intergenic
459 between *IER3* and *DDR1* (Fig.1c), only reached significance in European-ancestry populations
460 (P_{random}=2.6x10^{-10}) and was not correlated (r^2=0) with any of the SNPs reported under the Latino
461 admixture peak. The 6p21.33 region is gene rich, with many genes whose transcripts are
462 associated with our asthma signals, including *IER3*, *FLOT1*, *VARS2*, *MICB, HLA-C, MICA* in
463 blood[12-15,26] and *CDSN, TUBB, HLA-A*, *AGPAT1, HLA-C* and *MICA* in lung[14,16]
464 (Supplementary Table 14). Of these genes, *MICA*, *MICB* and HLA-Class I genes are known to
465 be involved in immune-related mechanisms. This 6p21.33 locus is more than 600 kb apart from
466 the previously reported 6p21.32 locus that spans HLA-Class II genes. Intensive sequencing
467 efforts and imputation of HLA classical alleles, which requires access to the raw genotype data,
468 are needed to dissect the whole HLA region. The second association at a known locus was at
469 10p14 (Fig.2d), where a GWAS in Japanese[9] reported association with adult asthma (lead SNP

rs10508372). We detected a new signal that reached significance only in the multi-ancestry meta-analysis (rs2589561; $P_{random}=3.5\times10^{-9}$) and was not correlated with rs10508372 in both European-ancestry and Japanese populations. The SNP rs2589561 is in a gene desert, 929 kb distal of *GATA3*. By using recently published promoter capture Hi-C data in hematopoietic cells[37], we found that two proxies of rs2589561 ($r^2>0.9$) lie in a region interacting with *GATA3* promoter, especially in immune CD4+T cells. This result and the function of GATA3 protein, a transcription factor that is a master regulator of differentiation of Th2 cells and innate lymphoid cells type 2[38], make *GATA3* an excellent candidate for asthma risk.

Asthma signals previously reported for asthma plus hay fever

Loci on chromosomes 8q21.13 and 16p13.13 were previously associated with the comorbid phenotype of asthma plus hay fever but not with asthma alone in one European-ancestry study[10]. In our study, the 8q21.13 lead SNP rs12543811 ($P_{random}=1.1\text{x}10^{-10}$ in the multi-ancestry analysis) lies between *TPD52* and *ZBTB10* (Fig.2c) and is in strong LD ($r^2=0.79$) with the previously reported asthma/hay fever SNP rs7009110. These two SNPs represent the same signal, as the association with rs12543811 becomes non-significant after conditioning on rs7009110. Thus, the 8q21.13 locus is likely to be implicated in allergic asthma. A functional analysis of the asthma/hay fever locus pinpointed *PAG1* as a promising candidate[39]. The chromosome 16p13.13 SNP rs17806299 is within an intron of *CLEC16A* (Fig.1e; $P_{random}=2.1\text{x}10^{-10}$ and $2.7\text{x}10^{-10}$ in European-ancestry and multi-ancestry meta-analyses) and in moderate LD ($r^2=0.66$) with the previously reported asthma/hay fever signal (rs62026376 in *CLEC16A*)[10]. However, the association of asthma with rs17806299 was completely removed after conditioning on rs12935657, a SNP present in our data and having $r^2=0.96$ with rs62026376, indicating that these SNPs represent the same signal and 16p13.13 is probably also an allergic asthma locus. The SNP rs17806299 is strongly associated with the expression of a nearby gene, *DEXI* in blood[13,26]. Similar observations of associations of *CLEC16A* SNPs with

autoimmune diseases and expression of *DEXI* together with chromosome conformation capture experiments implicated *DEXI* as the most likely candidate gene for autoimmune diseases[40]. The potential relevance of *DEXI* in allergic diseases has also been previously discussed[10].

It is of note that all the lead SNPs at the nine new asthma-associated loci lie in non-coding regions and were not found in linkage disequilibrium (LD; $r^2 \geq 0.8$) with missense variants.

Overlap of asthma associations with susceptibility loci for other phenotypes

We next explored whether the nine loci that harbored new signals for asthma *per se* overlapped with GWAS loci reported for allergy-related phenotypes, lung function phenotypes, or other immune-related diseases using the GWAS catalog[3]. Six of these nine asthma loci showed overlapping associations with allergy-related phenotypes and eight of them with auto-immune diseases and infection-related phenotypes (Table 2). The 6p21.33 locus which harbors HLA-Class I genes, had the greatest number of overlapping associations, which reflects the density of immune genes in that region. Moreover, three asthma loci overlapped with associations with lung function phenotypes.

We then expanded our search of overlap between the asthma association signals having multi-ancestry $P_{random} < 10^{-4}$ in this study and GWAS signals with all phenotypes and diseases in the GWAS catalog[3]. We examined 4,231 unique trait-loci combinations after applying a filtering process (Online Methods). We used the disease classification from Wang *et al.*[41] to group traits in the GWAS catalog. We summarized the overlap of GWAS catalog signals with asthma signals by the proportion of catalog SNPs having asthma *P*-values smaller than 10^{-4} in our analysis. This revealed significant overlap with autoimmune disease (10%, i.e. 49 out of 480 catalog SNPs show evidence for asthma association), consistent with the hypothesized shared susceptibility[42], moderate overlap with diseases having an inflammatory component (cardiovascular diseases, cancers, neuro-psychiatric diseases), and small to no overlap with

other diseases (Table 3). When investigating specific diseases and traits (Supplementary Table 15), the most significant overlap is with allergic phenotypes including atopic dermatitis. There is little to no overlap with other phenotypes that appear most frequent in the GWAS catalog (for example, there are no common associations with type 2 diabetes, Supplementary Table 15).

When we examined an even larger set of SNPs in the GWAS catalog (those that show an association with asthma at $P_{random} \leq 10^{-3}$ in our multi-ancestry meta-analysis), additional interesting pleiotropic signals emerged (Supplementary Table 16). In this analysis, we used a less stringent significance threshold to provide a larger set of candidate genes for future asthma studies and for pleiotropy analysis of asthma and diseases sharing genetic associations. This larger set of associations paints a more complete picture of asthma risk, revealing a wide range of pleiotropic effects for a number of traits, ranging from lung cancer and multiple sclerosis with rs3817963 in *BTNL2*, to schizophrenia and autism spectrum disorder with rs7914558 in *CNNM2*. Finally, this analysis generated an extended set of asthma candidate genes. Indeed, there are 210 SNPs in the GWAS catalog that are associated with asthma in our dataset at this association threshold of 10^{-3}. We estimated that the proportion of false positives among these is smaller than 1%. The most interesting signals are those with SNPs implicating novel pleiotropic asthma genes; for example, rs1464510 ($P_{random}=2.2 \times 10^{-5}$) in the *LPP* gene at 3q28 that was previously implicated in vitiligo and celiac disease.

Relationship between asthma associated genetic variants and epigenetic marks of gene regulation

Because nearly all lead SNPs at the 18 loci identified by this study lie in non-coding sequences, except for the missense variant (rs20541) within *IL13*, we investigated whether the asthma-associated variants and their proxies ($r^2 \geq 0.80$) were concentrated in cis-regulatory DNA elements to get greater insight into their potential functional role. We interrogated the integrated

functional data from 111 ROADMAP and 16 ENCODE reference epigenomes in a wide range of human cell types[27]. We focused on histone marks characterizing enhancers and promoters assayed in all 127 epigenomes and DNase I hypersensitivity sites available in 51 cell types. To assess enrichment of the asthma risk variants for co-localization with these regulatory elements, we used the Uncovering Enrichment through Simulation pipeline[43]. This approach first consists of generating random sets of SNPs, from the SNP panel used for analysis, that match to the characteristics of the original asthma-associated SNPs (distance from the nearest transcription start site, number of LD partners, minor allele frequency). The SNP data (original and random sets of SNPs and their LD partners) are then intersected with the cell-specific epigenome tracks of regulatory elements using IntersectBed[44] to determine which SNPs co-localize with a given type of regulatory elements in a track. For each track, the empirical P-value for enrichment is equal to the number of instances when the frequency of co-localization of the random SNP sets with a regulatory feature is greater than or equal to the frequency of the co-localization of the original asthma-associated SNP set with that feature divided by the number of random SNPs sets generated (here, 10,000). Benjamini-Hochberg false discovery rates (FDRs) were computed from these empirical P-values to correct for multiple testing (Online Methods). Only 16 out the 18 identified asthma loci were explored because we excluded the two loci spanning the HLA region due to the large amount of variability and extensive LD in this region. While asthma-associated variants were strongly enriched for co-localization with enhancer marks, there was only weak enrichment in promoter marks (Table 4 and Supplementary Table 17). This enrichment was slightly more pronounced when active enhancer states were examined separately from all active and inactive enhancer states. This enrichment was highest in leukocytes (27 leukocytes of which 19 (70%) are lymphocytes and monocytes). For example, a FDR≤5% for enrichment of asthma loci in active enhancers was observed in 100% of leukocytes compared to 50% of all cell types, and similarly for enrichment in both active and

inactive enhancers (FDR≤ 5% in 89% of leukocytes and in 43% of all cell types) (Table 4). The enrichment of asthma risk variants for co-localization with DNAse I hypersensitivity sites was intermediate between the enrichments in promoters and enhancers and was again increased in blood cells (FDR≤5% in 40% of leukocytes and 12% of all cell types) (Table 4 and Supplementary Table 18). The strong enrichment of asthma loci in enhancer marks, especially in immune cells, indicates that the associated genetic variants are likely involved in regulation of immune-related functions. This also suggests that epigenetic mechanisms may be key to promoting asthma, as evidenced for IgE levels, an asthma-associated phenotype[45].

Connectivity between asthma-associated loci

To characterize the degree of connectivity between genes located at the 18 asthma-associated loci, we applied the Gene Relationships Across Implicated Loci (GRAIL) text-mining approach[28]. Eleven genes showed significant similarity to genes in other associated loci at $P_{GRAIL}<0.05$ (7 of them being highly connected with $P_{GRAIL}<10^{-3}$) (Fig.3 and Supplementary Table 19). These genes were connected by keywords such as 'asthma', 'allergy', 'atopic', 'interleukin', 'cytokines', 'airway', and 'inflammation', which confirms the central role of immune-related mechanisms accounting for these connections.

Discussion

We present here the largest genome-wide association study of asthma in subjects of European-ancestry and in ancestry-diverse subjects. We discovered nine novel loci influencing asthma risk, using a stringent threshold of 10^{-8} for genome-wide significance. This study confirms that immune-related mechanisms are prominent in the etiology of asthma, but also brings novel insights that open new routes for future asthma research. Indeed, our data show that all the asthma-associated loci identified by this study are enriched in enhancer marks and are thus

likely to be involved in gene regulation. This study also highlights that the best candidates at a number of loci are involved in immune response to viruses or bacteria, which underlines the importance of infections in asthma risk. It further provides evidence for overlap of asthma loci with loci underlying other diseases that have an inflammatory component, such as cardiovascular diseases, cancers and neuro-psychiatric diseases. These collective findings both expand our current understanding of the genetic architecture of asthma as well as provide a more comprehensive resource for future genetic and epigenetic studies of asthma and further investigation of the mechanisms underlying the pleiotropic effects of genetic variants on asthma and a broad range of multifactorial diseases.

This study, which doubles the number of cases from the previous largest genome-wide studies[23,24], identified 613 SNPs that correspond to 22 distinct association signals at 18 loci that met criteria for genome-wide significance in European-ancestry or multi-ancestry populations. Pooling data from ethnically-diverse populations can increase power to detect new loci (four loci reached the genome-wide threshold only in the multi-ancestry analysis) but may also increase heterogeneity (two loci were only detected in European-ancestry populations). If we assume an asthma prevalence of 10%, the variance in liability to asthma explained by the 22 genome-wide significant variants of this study was estimated to be 3.5% (95% Confidence Interval: 2.0%-5.4%) of which 72% was accounted for by the known loci and 28% by the new loci.

The overall relative paucity of asthma risk loci detected by large-scale GWAS compared to other common diseases may be due to the clinical heterogeneity of asthma and the important role of differing environmental exposures. It is recognized that asthma is not a single disease but that the syndrome encompasses consistent groupings of various characteristics[46], including age of asthma onset, the severity of disease, occupational exposures and the varying response to treatment. Other subtypes may be defined by the character of the inflammatory infiltrate

(eosinophilic or neutrophilic). It is thus possible that additional asthma loci will be revealed by studies targeting more specific asthma sub-phenotypes and/or taking into account environmental exposures.

The enrichment of the asthma loci identified by this study in enhancer marks, mainly in lymphocytes and monocytes, suggests that the underlying causal variants may modify the gene regulatory landscape in immune cells. However, known asthma genes (*IL1RL1, TSLP, IL33, ORMDL3/GSDMB*) are expressed in the airway epithelium where they contribute to modulating airway inflammation. Investigation of epigenetic marks in airway epithelial cells may bring additional insight. The involvement of several genes (*BACH2*, *NDFIP1* or genes within HLA), at the identified loci, in immune response to viruses provides additional evidence for a potential modulating effect of asthma-associated SNPs on the association of asthma with respiratory viral infections, as reported at the 17q12-21 locus[47-49]. The overlap of asthma-associated loci with those underlying a broad range of diseases strengthens the growing importance of pleiotropy in multifactorial diseases.

In conclusion, future discoveries might come by exploring more complex models of asthma phenotypes and through the joint analysis of asthma and other immune-mediated and inflammatory diseases. The central role of gene regulatory mechanisms highlighted by our study might prompt genome-wide explorations of the epigenome while integrating information on genetic variation and environmental exposure histories.

URLs

National Human Genome Research Institute (NHGRI) and European Bioinformatics Institute (EBI) catalog of published genome-wide association, https://www.ebi.ac.uk/gwas/; 1000 Genomes Project Consortium Phase 3, http://www.internationalgenome.org/; Genome-wide Complex Trait Analysis (GCTA), http://www.complextraitgenomics.com/software/gcta;

645 Blood-eQTL database, http://genenetwork.nl/bloodeqtlbrowser/; GTEx,

646 http://www.gtexportal.org/; MuTHER database, http://www.muther.ac.uk/; eQTL database in

647 lymphoblastoïd cell lines from MRCA and MRCE families,

648 https://www.hsph.harvard.edu/liming-liang/software/eqtl/; GHS_Express,

649 http://genecanvas.ecgene.net/uploads; ROADMAP and ENDCODE epigenomics data,

650 http://egg2.wustl.edu/roadmap/web_portal/; Uncovering Enrichment through Simulation

651 (UES) pipeline, https://github.com/JamesHayes/uesEnrichment; GRAIl,,

652 https://www.broadinstitute.org/ mpg/grail/; VIZGRAIL,

653 http://software.broadinstitute.org/mpg/grail/vizgrail.html, LocusZoom,

654 http://locuszoom.sph.umich.edu/ locuszoom/

655

ONLINE METHODS

GWAS Studies and Data Shared

All 66 genome-wide association studies that form the TAGC consortium are described in the Supplementary Note and summarized in Supplementary Table 1. These studies included 56 studies of European-ancestry (19,954 cases, 107,715 controls), seven studies of African-ancestry (2,149 cases, 6,055 controls), two Japanese studies (1,239 cases, 3,976 controls) and one Latino study (606 cases, 792 controls), making a total of 23,948 cases and 118,538 controls. There were 27 studies including only childhood-onset asthma (defined as asthma diagnosed at or before 16 years of age) which allowed us analyzing separately a pediatric subgroup (8,976 cases, 18,399 controls). All subjects provided informed consent to participate in genetic studies and local ethics committees for each of the individual studies approved the study protocol. Definition of asthma was based on doctor's diagnosis and/or standardized questionnaires (see Supplementary Note for details). The samples were genotyped on a variety of commercial arrays, detailed in the Supplementary Note and Supplementary Table 2. GWAS were performed on imputed SNP data that were generated using Hapmap2 as reference panel and one of the commonly used imputation software (Supplementary Note and Supplementary Table 2). In each dataset, the effect of each individual SNP on asthma, assuming an additive genetic model, was estimated through a logistic regression-based approach and expressed in terms of a regression coefficient with its standard error; the detailed methodology and software used for analysis by each study can be found in the Supplementary Note and Supplementary Table 2.

Imputation, quality control (including adjustments for population stratification) and analysis was done by each group independently and data on a predefined set of 3,952,683 autosomal SNPs was shared. These SNPs were those of the HapMap Phase 2, release 21 panel in subjects from European, Asian and African-ancestry that were filtered using SNP annotation from the

682 build 37.3 of the reference sequence and dbSNP b135 (31,587 SNPs (0.8% of all SNPs) from
683 previous annotations that showed discrepancies with the chosen annotation were deleted). The
684 variables that were shared contained the study name, general information on SNPs (rs number,
685 chromosome, position, alleles (baseline and effect alleles as used in the analysis by each study),
686 SNP status (imputed or genotyped SNP and whether the SNP genotype or imputed value was
687 used in computation), quality control (QC) indicators (call rate and *P*-value for Hardy-
688 Weinberg (HW) equilibrium test for genotyped SNPs, software used for imputation and
689 imputation quality score for imputed SNPs), allele frequencies in cases and controls and
690 information on association statistics (regression coefficient for SNP effect, standard error of
691 regression coefficient, Z scores, *P*-values associated with Z score statistic). Association
692 statistics for at least one SNP in at least one study were available for 2.93 million SNPs.
693
694 **Quality control of shared data**
695 For each SNP, the alleles on the Hapmap2 template (reference and alternate alleles on the
696 positive strand) were compared to the alleles (baseline and effect alleles) used in the analysis
697 by each group. When necessary, the association variables (allele frequencies, regression
698 coefficient for SNP effect, Z score) were swapped to match the reference/alternate alleles of the
699 template. Data for each SNP showing any ambiguity or error in assignment to the template were
700 set to missing. In addition, a number of QC checks were done regarding the name, format, range
701 of possible values for all shared variables mentioned in the previous paragraph as well as
702 consistency checks across variables. Any problem or inconsistency was corrected, otherwise
703 the data for that SNP were set to missing. Strict QC criteria were used for inclusion of a SNP
704 in the analysis. When a SNP genotype was used in the study analysis, these criteria were: call
705 rate $\geq$97%, *P*-value for HW test $\geq 10^{-6}$ and minor allele frequency (MAF) $\geq$0.01 in both controls
706 and cases. When a SNP imputed value was used in the analysis, the criteria were: imputation

quality score ≥0.5 and MAF≥0.01 in both controls and cases. The distribution of the summary statistics (regression coefficient for SNP effect, standard error, Z score) of all SNPs passing QC was examined for each study; SNPs that still showed extreme Z scores (≥ 7 or ≤-7) after QC were excluded.

Meta-analysis of GWAS

We used both inverse variance fixed-effects and random-effects models when the meta-analyses included a large number of studies (European-ancestry, multi-ancestry and pediatric sub-group meta-analyses), which allows an accurate estimate of the between-study variance based on the DerSimonian and Laird method[50]. We used a fixed-effects model for the meta-analyses of the African-ancestry, Japanese and Latino populations, which included at most seven studies. For all these meta-analyses, we used the SNP regression coefficient and standard error from each study for which the SNP passed QC[11]. All meta-analyses were done with Stata® version 14.1 (STATA Corp., College Station, Texas, USA). To minimize the false-positive findings and to obtain robust results, we examined the combined results for SNPs for which at least two-thirds of the studies contributed to a meta-analysis. We applied a stringent threshold of P_{random} (or P_{fixed}) of 10^{-8} to declare a SNP effect as genome-wide significant. This threshold, which is stricter than the standard threshold of $5x10^{-8}$, classically used in GWAS, allows minimizing the family-wise error rate. We checked that using the $5x10^{-8}$ threshold did not change the number of asthma-associated loci but only the list of significant SNPs at a locus. For each lead SNP, we defined a support interval around the lead SNP designated as "locus"; the bounds of this interval were the positions of the two most extreme SNPs among all SNPs lying on each side of the lead SNP and having P_{random} (or P_{fixed}) $\leq 10^{-6}$ (i.e. SNPs with *P*-values within two orders of magnitude of the critical genome-wide significance threshold). In each meta-analysis, heterogeneity of per-SNP effect sizes across studies was assessed using the

Cochran's Q statistic[11]. Heterogeneity between the four ethnic groups (European-ancestry, African-ancestry, Japanese and Latino) was also tested using the Cochran's statistic applied to meta-analysed results obtained in each ethnic group and all groups combined.

Conditional analysis of asthma-associated loci

The Genome-wide Complex Trait Analysis (GCTA) software[25] (see URL section) was used to perform approximate conditional analysis for all loci with at least one SNP reaching the genome-wide significance level. This approximate conditional analysis is based on the summary meta-analysis statistics obtained under a fixed-effects model and takes into account the correlations among SNPs, that are estimated from a large reference population included in the meta-analysis. Approximate conditional analysis was only carried out in the European-ancestry ethnic group which can be assumed to share a similar LD pattern and represents the largest ancestry-specific dataset and the only one to show genome-wide significant results. As this analysis assumes no heterogeneity in SNP effect size across studies, loci showing evidence for heterogeneity ($P_{het} \leq 0.05$ based on the Cochran's Q test) for a large portion of the locus were excluded. This was the case of the 9p24.1 locus that shows evidence for heterogeneity across the whole locus and the 17q12-21 locus where there is evidence for heterogeneity across a 200 kb segment within that locus. However, there is no heterogeneity between studies at 17q12-21 in the pediatric sub-group while there is still heterogeneity at 9p24.1 in this sub-group. Therefore, GCTA was applied to all 18 asthma-associated loci, shown in Table 1, except for 9p24.1 and 17q12-21 loci, using summary statistics from the meta-analysis of all European-ancestry studies. For the17q12-21 locus, this analysis was restricted to the European-ancestry pediatric sub-group. We used the large ECRHS (European Community Respiratory Health Survey) dataset as the reference sample to estimate LD. This dataset was genotyped using Illumina Human610Quad array, as part of the GABRIEL consortium, and included 2101

unrelated individuals after QC[23]. Imputation was done using the MACH software[51] and Hapmap2, release 21 panel; only well-imputed SNPs (imputation quality score rsq>0.8) and with minor allele frequency (MAF) ≥1% were kept in this reference panel. For each asthma-associated locus, the region explored by conditional analysis extended by 500 kb on each side of the two extreme SNPs defining the support interval around the lead SNP (see preceding paragraph). However, we reduced that extension to 250 kb for the 6p21.33 and 6p21.32 loci to avoid overlap. The length of the regions explored by conditional analysis varied from 1.01 Mb to 1.63 Mb. Within each investigated region by conditional analysis, fixed-effects summary meta-analysis data for SNPs belonging to that region were adjusted for the lead SNP using the --cojo-cond option. If there was an additional SNP meeting the Bonferroni-corrected threshold for the total number of SNPs overall all regions investigated by GCTA (P=4.1x10^{-6}) after adjustment for the lead SNP, we performed an additional round including both SNPs. If the remaining SNPs had P-values greater than 4.1x10^{-6}, no further analysis was performed. The results of this analysis are reported in Supplementary Table 13.

Identification of cis-eQTLs at new asthma-associated loci

To gain a better understanding of the potential genes driving the association signals at the novel asthma loci, we defined a list of SNPs to be interrogated that included the lead SNPs, the secondary signals identified by conditional analysis and all SNPs in LD with these SNPs (r^2 comprised between 0.5 and 1). To search for cis-expression quantitative trait loci (eQTLs) within at most 1 Mb of each investigated SNP, we interrogated six publically available eQTL databases by giving priority to cell types more likely to be involved in asthma biology (blood cell types and lung tissue): (i) a meta-analysis of the transcriptional profiles from peripheral blood cells of 5,311 Europeans (the blood eQTL browser[13]);(ii) the gene expression data of 777 lymphoblastoïd cell lines (LCLs) from the MuTHER database[12]; (iii) the transcription profiles

of 405 and 550 lymphoblastoïd cell lines from UK asthma (MRCA) and eczema (MRCE) family members, respectively[15]; (iv) the eQTL data from monocytes of 1490 subjects included in the GH-express database[26]; (v) the GTEx eQTL Browser with data from multiple tissues including blood and lung[14]; (vi) the transcriptional profiles from lung tissues of 1111 subjects[16].

Identification of missense variants at asthma-associated loci

To complement the eQTL analysis, we searched whether the lead asthma-associated SNPs and secondary signals were in strong LD ($r^2 \geq 0.8$) with missense variants using the HaploReg V4.1 tool (see URL section).

Overlap of asthma associations with susceptibility loci for other phenotypes

Overlap of novel genome-wide significant asthma loci with associations with allergy-related phenotypes/diseases and immune-related diseases as well as lung function phenotypes was first annotated using the March 24, 2015 version of the NHGRI-EBI (National Human Genome Research Institute and European Bioinformatics Institute) GWAS catalog[3] downloaded from https://www.ebi.ac.uk/gwas/. We then used this catalog to systematically investigate the overlap of asthma signals having $P_{random} \leq 10^{-4}$ in the multi-ancestry meta-analysis with association signals of all diseases and traits in the catalog. That version of the catalog had 19,080 SNP entries, and 16,047 of those SNPs had a TAGC asthma association *P*-value. To investigate pleiotropy, we filtered out SNPs associated with asthma in the database, SNPs that have a reported GWAS *P*-value larger than 10^{-7} (with the intent of removing some of the potential false positives in the catalog) and SNPs that are duplicated (i.e., remove disease-SNP duplications). This reduced the number of entries to 5,927. Note that this process did not remove SNPs in perfect LD associated with the same disease, nor SNPs that were present multiple times in the database as associated with different phenotypes. For some diseases or quantitative traits,

there were multiple SNPs in the same region reported in the catalog potentially yielding redundant information. Some of the SNPs could be in strong LD, but some could reflect independent signals. To avoid possible duplication of signals, we decided to keep only unique trait-loci combinations as reflected by the variables "Disease.Trait" and "Region" in the catalog. There were 4,231 unique entries left after this filtering step. Diseases/traits in the GWAS catalog were grouped using the classification from Wang *et al*. [41] We summarized the overlap of GWAS catalog signals with asthma signals by the proportion of catalog SNPs with asthma *P*-values smaller than 10^{-4} in our analysis. The significance of overlap was estimated by the binomial tail probability for observing the number of TAGC SNPs with $P_{random} \leq 10^{-4}$ among the number of SNPs reported in the GWAS catalog for a group of diseases. The significance threshold for enrichment in shared associations between a disease group and asthma was set equal to 0.05 divided by the number of disease groups investigated using a Bonferroni correction. Finally, we examined a larger set of SNPs in the GWAS catalog that show an association with asthma at $P_{random} \leq 10^{-3}$ in TAGC multi-ancestry meta-analysis and estimated the proportion of false positives among those SNPs.

Relationships between asthma associated genetic variants and epigenetic marks of gene regulation

To get better insight into the functional role of the genetic variants at the novel and known asthma loci identified by this study, we investigated whether the lead SNPs and their proxies ($r^2 \geq 0.80$) were concentrated in cis-regulatory DNA elements. We used the Uncovering Enrichment through Simulation pipeline[43] that was adapted to the current study. This approach tests if GWAS-identified SNPs are enriched in particular functional annotations through use of Monte Carlo simulations. The original set of asthma-associated SNPs included the lead SNPs

at each locus (ie one SNP per asthma-associated locus, as recommended by Hayes *et al*[43]). We excluded the two associated loci that span the HLA region (6p21.33 and 6p21.32) because of the high amount of variability and LD in this region. Each of the original lead SNPs is categorized by its distance from the nearest transcription start site (TSS) and number of LD partners ($r^2 \geq 0.8$). Quartiles for both the TSS distance and LD partner count are calculated and the initial SNPs are binned accordingly. Then, SNPs from the whole set of imputed SNPs used for analysis are binned according to the original SNP criteria (distance from the closest TSS, number of LD partners, and also MAF). Random SNP sets are chosen, matching to the original bin frequencies. LD partners ($r^2 > 0.8$) for both the original lead SNPs and random SNPs are retrieved. The SNP data, including the original and random sets of SNPs and their corresponding LD partners ($r^2 \geq 0.8$), are intersected with the cell-specific epigenome tracks of regulatory elements using the BedTool's intersectBed[44], to determine which SNPs co-localize with a given type of regulatory elements (for example, enhancers or promoters) in a track. For each specific track, we computed an empirical-*P* value using 10,000 random SNP sets (this *P*-value is equal to r_{loci}/n where r_{loci} is the number of instances when the frequency of co-localization of the random SNP sets with the regulatory feature is greater than or equal to the frequency of co-localization of the original SNP set with that feature and n is the number of random SNP sets chosen (here, 10,000). We used the Benjamini-Hochberg false discovery rates (FDR) to correct for multiple testing. We interrogated the functional data from 111 ROADMAP reference epigenomes and 16 additional epigenomes from ENCODE (Encyclopedia of DNA elements) that are available in a wide range of human cell and tissue types[27] (data downloaded from http://egg2.wustl.edu/roadmap/data/). We focused on enhancers and promoters that were defined using the 15-state model trained from the core set of five chromatin marks (H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3) assayed in all 127 epigenomes. We also examined enrichment in DNase I hypersensitivity sites that are available in 51 cell types.

Connectivity between asthma-associated loci

We used GRAIL (Gene Relationships Across Implicated Loci)[28] to assess the relatedness between asthma associated loci. As described in detail previously[28], to define the genes near each SNP, GRAIL finds the furthest neighboring SNPs in the 3′ and 5′ direction that are in LD (r^2>0.5) and proceeds outward in each direction to the nearest recombination hotspot. All genes that overlap that interval are considered implicated by the SNP. If there are no genes in that region, the interval is extended by 250 kb in either direction. We took the lead SNPs at all asthma-associated loci identified by this study as a seed and queried loci to investigate biological connectivity among those loci. The relatedness between genes belonging to these loci was determined through text-mining of PubMed abstracts. Each individual gene at each locus was tested for significant enrichment in GRAIL connectivity to genes located at the other loci. The relationships between genes belonging to different loci that met the nominal threshold $P_{GRAIL} < 0.05$ were visualized using VIZGRAIL[52].

Variance explained by the asthma associated genetic variants

We estimated the variance in liability to asthma explained by the 22 distinct genome-wide significant SNPs (18 lead SNPs plus four secondary signals identified by approximate conditional analysis) at the 18 asthma-associated loci using a method based on the liability threshold model[53] and assuming a prevalence of asthma of 10%. The variance in liability to asthma explained by individual SNPs was summed over all 22 significant variants. For the loci that included two SNPs (lead SNP and secondary signal), we used the SNP effect sizes estimated by approximate joint analysis using GCTA[25]. We also estimated the variance in liability to asthma explained by the nine lead SNPs at the nine new asthma loci and by the 13 distinct genome-wide significant signals at the nine known loci.

Data availability

At time of submission, the summary statistics for association of SNPs with asthma that support the findings of this study are available to the Editor and reviewers from the corresponding author upon request. The summary statistics of the meta-analysis that support the findings of this study will be made publically available through a link from the GWAS entry catalog for the TAGC study on the EBI (European Bioinformatics Institute) web site (https://www.ebi.ac.uk/gwas/downloads/summary-statistics) as soon as the paper is accepted for publication.

Acknowledgments

Individual study acknowledgment and funding statements can be found in the Supplementary Information.

Author Contributions

All authors provided critical review of the manuscript.

Study management: F.D., K.C.B., W.O.C.C., M.F.M., C.O. and D.L.N.
F.D. and D.L.N. designed the study, designed and carried out the statistical analysis and wrote the manuscript. K.C.B., W.O.C.C., M.F.M. and C.O. designed the study and wrote the manuscript. P.M-J., M.B., A.V., S.L. and H.M. carried out the quality control of the data and performed statistical analysis.

AAGC (Australia): Study PI: M.A.F., M.C.M., C.F.R and P.J.T.; Data collection or analysis: M.A.F., M.C.M., C.F.R., G.J. and P.J.T.

ALLERGEN CAPPS and SAGE: Study-PI: A.B., M.C-Y., D.D. and A.L.K.; Data collection or analysis: D.D. and J.E.P.; Study Phenotyping: A.B., M.C-Y.; **SLSJ:** Study PI: T.J.H. and C.L.

APCAT: Study PI: J.H.N., M.R.J., G.T.O., N.J.W., V.S., M.H., P.J., J.G.E. and O.T.R.; Data collection or analysis: A.C.A., S.V., J.H.Z., M.K., E.W., T.L., A.R., Z.G., R.S., T.L., T.H., A.P., M.K.

CARe Consortium : Study-PI: J.N.H., S.S.R.; Data collection or analysis: C.D.P, D.B.K, L.J.S., R.K., K.M.., W.B.W.

MESA: Study PI: R.G.B, S.S.R; Data collection or analysis: K.M.D, A.M.

ARIC: Study PI: S.J.L.; Data collection or analysis: S.J.L., L.R.L., B.R.J.

deCode Genetics: Study PI: D.F.G., I.J., K.S., U.T., G.T.; Data collection or analysis: I.J., D.F.G. and G.T.; Study Phenotyping: U.S.B.

EAGLE Consortium: PI: HB. **COPSAC**: Study PI: H.B; Data analysis: E.K. and J.W.; Study Phenotyping: K.B.; **DNBS:** Study PI: M.M.; Data collection or analysis: B.F and F.G.; **GINIplus/LISAplus:** Study PI:J.H.; Data collection or analysis: M.S. and C.M.T.T.; Study Phenotyping: J.H.; **MAAS:** Study PI: A.S. and AC.; Data collection or analysis: J.A.C.; **RAINE:** Study PI: P.H.; Data collection or analysis: W.A. and C.E.P.; **GENERATION R:** Study PI: J.C.dJ.; Data collection or analysis: R.J.P.vV., L.D. and V.W.V.J.

B58C: PI and Statistical Analysis: D.P.S.

EVE: Study-PI: C.O., D.L.N., K.C.B., E.B., E.B., J.G., F.G., S.J.L., F.M., D.M., I.R., S.T.W., L.K.W., B.AR.; Data collection or analysis: D.L.N., J.G., S.J.L., D.M., D.T., B.A.R., B.E.H., P.G., M.T.S., C.E., B.R.N., J.J.Y., A.M.L., R.M.

JAARC: Study PI: T.T.; Data collection or analysis: T.T., A.T. and M.K.

JPAC: Study PI: E.N.; Data collection or analysis: H.H. and K.M.

GABRIEL CONSORTIUM: Study PI: W.O.C.C. and E.V.M.; Genotyping: M.L.; Data Analysis: E.B., F.D., M.F. and D.P.S. **EGEA:** Study PI: V.S. and F.D.; Genotyping, data collection or analysis: M.L. and E.B.; **ALSPAC:** Study PI: J.H.; Data collection or analysis: W.L.M. and R.G.; Study phenotyping: J.H.; **ECRHS:** Study PI: D.J.; Data collection or analysis: C.J. and J.H.; **BAMSE:** Study PI: EM, M.W. and G.P.; **BUSSELTON:** Study PI: A.W.M., J.B. and A.J.; Data collection or analysis: J.H., A.W.M. and A.J.; Study Phenotyping: A.W.M. and A.J.; **GABRIEL Advanced Studies;** Study PI: E.V.M.; Data collection or analysis: M.K. and J.G.; **KSMU:** Study PI: A.P.; Data collection or analysis: M.S. and V.I.; **MRCA-UKC**: Study PI: W.O.C.C. and M.M.; Data collection or analysis: L.L.; **MAGICS:** Study PI: MK; Data collection or analysis: A.vB. and S.M.; **MAS:** Study PI: Y.L.; Data collection or analysis: S.L. and I.M.; **PIAMA:** Study PI: G.H.K. and D.S.P.; Data collection or analysis: U.G.; **SAPALDIA:** Study PI: N.P.H.; Data collection or analysis: M.I. and A.K. **TOMSK:** Study PI: L.O. and V.P.; Data collection or analysis: M.F. and P.S.; **UFA:** Study PI: E.K.; Data collection or analysis: A.K. and Y.F. **INDUSTRIAL:** Study PI: D.H. and T.S.; Data collection or analysis: I.M.W. and V.S.; **SEVERE**: Study PI: A.B., K.F.C. and C.E.B.

NTR: Study PI: D.I.B.; Data collection or analysis: J.J.H., H.M. and G.W.

Rotterdam: Study PI: A.H., B.H.S., G.G.B.; Data collection or analysis: G.G.B., B.H.S., D.W.L., L.L., and A.G.U.

DAGC: Study PI: G.H.K. and D.S.P.; Data collection or analysis: J.A., M.A.E.N. and J.M.V.

Competing Financial interests

The authors who are affiliated with deCODE (D.F.G., I.J., K.S., U.T., G.T. and U.S.B) are employees of deCODE genetics/AMGEN. All other co-authors did not declare any conflict of interest.

Table 1 Genetic loci associated with asthma in European-ancestry and multi-ancestry meta-analyses

						European-ancestry meta-analysis				Multi-ancestry meta-analysis			
Region[a]	No. Sig SNPs Eur-anc/ Multi-anc[b]	Lead SNP[c]	Position[d]	Nearby genes[e]	Allele[f] (R/E)	EAF[g]	OR (95%CI)[h]	P_{random}[i]	P_{het}[j]	OR (95%CI)[h]	P_{random}[i]	P_{het}[j]	P_{ethnic}[k]
New asthma susceptibility loci													
5q31.3	0 / 2	rs7705042	141,492,419	*NDFIP1,GNDPA1,SPRY4*	C/A	0.63	1.08 (1.05-1.11)	1.6×10^{-6}	0.07	1.09 (1.06-1.12)	7.9×10^{-9}	0.11	0.53
6p22.1	4 / 0	rs1233578	28,712,247	*GPX5, TRIM27*	A/G	0.13	1.11 (1.07-1.15)	5.3×10^{-9}	0.82	1.09 (1.05-1.12)	5.9×10^{-7}	0.56	0.003
6q15	26 / 26	rs2325291	90,986,686	*BACH2,GJA10,MAP3K7*	G/A	0.33	0.91 (0.89-0.93)	8.6×10^{-13}	0.78	0.91 (0.89-0.94)	2.2×10^{-12}	0.8	0.39
12q13.3	0 / 1	rs167769	57,503,775	*STAT6,NAB2,LRP1*	C/T	0.40	1.08 (1.05-1.11)	1.6×10^{-7}	0.19	1.08 (1.05-1.11)	3.9×10^{-9}	0.31	0.87
17q21.33	2 / 1	rs17637472	47,461,433	*ZNF652,PHB*	G/A	0.39	1.08 (1.05-1.11)	3.3×10^{-9}	0.56	1.08 (1.05-1.11)	6.6×10^{-9}	0.35	0.12
New signals at loci previously associated with asthma in ancestry-specific populations													
6p21.33	34 / 34	rs3131064	30,763,893	*IER3,DDR1*	T/C	0.14	1.12 (1.08-1.16)	2.6×10^{-10}	0.54	1.10 (1.06-1.13)	2.6×10^{-8}	0.46	0.02
10p14	0 / 1	rs2589561	9,046,645	*GATA3,CELF2*	A/G	0.89	0.90 (0.87-0.94)	1.4×10^{-8}	0.78	0.91(0.88-0.94)	3.5×10^{-9}	0.82	0.25
Asthma signals previously reported for asthma plus hay fever													
8q21.13	0 / 16	rs12543811	81,278,885	*TPD52,ZBTB10*	G/A	0.66	0.93 (0.91-0.95)	3.4×10^{-8}	0.47	0.92 (0.9-0.95)	1.1×10^{-10}	0.54	0.24
16p13.13	9 / 9	rs17806299	11,199,980	*CLEC16A,DEXI,SOCS1*	G/A	0.20	0.90 (0.88-0.93)	2.1×10^{-10}	0.51	0.91 (0.88-0.94)	2.7×10^{-10}	0.49	0.58
Known asthma loci													
2q12	119 / 135	rs1420101	102,957,716	*IL1RL1,IL1RL2,IL18R1*	C/T	0.37	1.12 (1.10-1.15)	9.1×10^{-20}	0.63	1.12 (1.09-1.15)	3.9×10^{-21}	0.61	0.64
5q22.1	28 / 21	rs10455025	110,404,999	*SLC25A46,TSLP*	A/C	0.34	1.15 (1.12-1.18)	2.0×10^{-25}	0.53	1.15 (1.12-1.18)	9.4×10^{-26}	0.57	0.27
5q31	16 / 28	rs20541	131,995,964	*IL13,RAD50,IL4*	A/G	0.79	0.89 (0.86-0.91)	1.4×10^{-14}	0.73	0.89 (0.87-0.92)	5.0×10^{-16}	0.77	0.62
6p21.32	78 / 82	rs9272346	32,604,372	*HLA-DRB1,HLA-DQA1*	G/A	0.56	1.16 (1.13-1.19)	4.8×10^{-28}	0.46	1.16 (1.12-1.19)	5.7×10^{-24}	0.14	0.43
9p24.1	30 / 57	rs992969	6,209,697	*RANBP6,IL33*	A/G	0.75	0.85 (0.82-0.88)	1.1×10^{-17}	0.01	0.86 (0.83-0.88)	7.2×10^{-20}	0.02	0.53
11q13.5	3 / 4	rs7927894	76,301,316	*C11orf30, LRRC32*	C/T	0.37	1.10 (1.07-1.13)	3.5×10^{-11}	0.38	1.10 (1.08-1.13)	2.2×10^{-14}	0.56	0.48
15q22.2	3 / 10	rs11071558	61,069,421	*RORA,NARG2,VPS13C*	A/G	0.14	0.89 (0.85-0.92)	1.9×10^{-10}	0.44	0.89 (0.86-0.92)	1.3×10^{-9}	0.19	0.06
15q22.33	13 / 13	rs2033784	67,449,660	*SMAD3,SMAD6,AAGAB*	A/G	0.30	1.11 (1.08-1.14)	2.5×10^{-14}	0.75	1.10 (1.08-1.13)	7.4×10^{-15}	0.76	0.48
17q12-21	158 / 173	rs2952156	37,876,835	*ERBB2,PGAP3,C17orf37*	A/G	0.70	0.86 (0.84-0.88)	7.6×10^{-29}	0.55	0.87 (0.84-0.89)	2.2×10^{-30}	0.52	0.35

[a]Cytogenetic position of the genome-wide significant SNPs ($P_{random} \leq 10^{-8}$). [b]Number of genome-wide significant SNPs in the European-ancestry/multi-ancestry meta-analyses. [c] All lead SNPs shown in this table have imputation quality scores > 0.80 in ≥ 80% of studies; these SNPs are the lead SNP in the European-ancestry meta-analysis if genome-wide significant otherwise they are the lead SNPs in the multi-ancestry meta-analysis [d] Build 37. [e]The gene where eventually the SNP lies is first indicated followed by the previous gene and next gene. [f]R=reference allele /E=effect allele. [g]EAF=Effect allele frequency in European-ancestry populations (weighted average of allelic frequencies from individual studies). [h]Odds-ratios (ORs) were computed for the effect allele under a random-effects model. [i]*P* value for association of SNP with asthma (random-effects model). [j]*P* value for heterogeneity across studies using Cochran's Q test. [k]*P* value for heterogeneity across ethnic groups.

Table 2. Main characteristics of the nine loci harboring novel associations with asthma

Region[a]	# of distinct signals[b]	Location of lead SNP[c]	Cis-eQTLs in blood (B) and lung tissue (L)[d]	Association with allergy-related and lung function phenotypes[e]	Association with auto-immune diseases and other immune-related traits[e]
New asthma susceptibility loci					
5q31.3	1	*NDFIP1* (intron)	B: *NDFIP1* ($2.7x10^{-9}$)		IBD
6p22.1	1	intergenic		Lung function	
6q15	1	*BACH2* (intron)	B: *BACH2* ($2.9x10^{-10}$)		MS, T1D, CD, IBD, V, IGG
12q13.3	1	*STAT6* (intron)	B: *STAT6* ($9.8x10^{-198}$) L: *STAT6* ($3.7x10^{-37}$)	IgE (total, specific) Lung function	Pso, ISP_IFN
17q21.33	1	intergenic	B: *GNGT2* ($2.1x10^{-52}$)	Atopic dermatitis	ISP_IL2
New asthma signals at loci previously associated with asthma in ancestry-specific populations					
6p21.33	3	intergenic	B: *IER3* ($8.6x10^{-47}$), *FLOT1* ($9.1x10^{-33}$), *VARS2* ($3.7x10^{-30}$), *MICB* ($6.9x10^{-21}$), *HLA-C* ($4.1x10^{-18}$), *MICA* ($7.4x10^{-12}$), L: *CDSN* ($4.8x10^{-19}$), *TUBB* ($2.3x10^{-16}$), *HLA-A* ($1.9x10^{-13}$), *AGPAT1* ($2.4x10^{-12}$), *HLA-C* ($5.6x10^{-12}$), *MICA* ($2.5x10^{-9}$)	IgE (total, specific), Self-reported allergy, Atopic dermatitis, Lung Function	SLE, UC, RA, IBD, BS, GD, SS, AS, Pso, UC, V, WBC, MoC, DS, HIV-1, SJS, HB, HBV, IMN, CD4:CD8 ratio, HIV-1C
10p14	1	intergenic	None	Self-reported allergy	RA, ISP_IL1B, ISPV
Asthma signals previously reported for asthma plus hay fever					
8q21.13	1	intergenic	None	Atopic dermatitis, Asthma + hay fever Self-reported allergy	RA
16p13.13	1	*CLEC16A* (intron)	B: *DEXI* ($2x10^{-43}$)	Atopic dermatitis, Asthma + hay fever	T1D, PBC, MS, RA, IBD, CD, LEP

[a]Cytogenetic position of the genome-wide significant SNPs (*Prandom* $\leq 10^{-8}$).[b] Number of distinct association signals at a locus detected by conditional analysis using the Genome-wide Complex Trait Analysis (GCTA) software[25] (Online Methods). [c] The protein coding genes flanking intergenic SNPs are shown in Table 1. [d] Cis-genes whose expression is

associated with lead asthma-associated SNPs or proxies ($r^2 \geq 0.8$) using six eQTL databases from whole blood[13,14], lymphoblastoïd cell lines[12,15], monocytes[26] and lung[14,16]; only genes showing the strongest associations are shown in this table (Supplementary Table 14 for details). [e]Associations with other diseases are based on the GWAS catalog[3]: IBD=Inflammatory bowel diseases (Crohn's disease), MS=multiple sclerosis, , T1D=type 1 diabetes, CD=celiac disease, V=vitiligo, IGG=IgG Glycosylation, Pso=psoriasis, ISP_IFN=Immune Response to Smallpox (secreted IFN-alpha), ISP_IL2 Immune Response to Smallpox (secreted IL2), SLE=Systemic Lupus Erythematosus, UC=Ulcerative colitis, RA=Rheumatoid arthritis, BS=Behcet syndrome, GD=Grave's disease, SS=Systemic sclerosis, AS=Ankylosing spondylitis, WBC=White Blood cell count, MoC=monocyte count, DS=Dengue shock, HIV-1=HIV-1-susceptibility, SJS=Stevens-Johnson syndrome, HB=Hepatitis B infection, HBV=Hepatitis B vaccine response, IMN=Idiopathic membranous nephropathy, CD4:CD8=CD4:CD8 lymphocyte ratio, HIV-1C=HIV-1 control, ISP_IL1B=Immune Response to Smallpox (secreted IL-1 beta), ISPV=Immune response to smallpox vaccine (IL-6), PBC=Primary biliary cirrhosis, LEP=Leprosy.

Table 3. Overlap of TAGC asthma-associated SNPs[a] with GWAS catalog association signals by disease group

Disease Group[b]	Number of GWAS catalog association signals	Number of SNPs associated with asthma at $P_{random} \leq 10^{-4}$ in the multi-ancestry meta-analysis	Enrichment *P*-value[c]
Cardiovascular	743	20	7.8×10^{-42}
Body size and morphology	346	2	5×10^{-4}
Immune/Autoimmune	480	49	3.0×10^{-129}
Nervous system	242	4	1.4×10^{-8}
Blood	594	10	1.3×10^{-19}
Neuropsychiatric	114	5	1.5×10^{-12}
Cancer	417	7	4.0×10^{-14}
Endocrine system	276	2	4×10^{-4}
Digestive system	347	16	1.4×10^{-37}
Eyes	177	2	2×10^{-4}
Respiratory system	85	2	3.6×10^{-5}
Infectious disease/Infection	104	2	5.3×10^{-5}
Urinary system	144	1	0.015
Alcohol, smoking, and illicit substances	30	0	1
Musculoskeletal system	132	0	1

[a] TAGC asthma-associated SNPs are those having $P_{random} \leq 10^{-4}$ in the multi-ancestry meta-analysis. [b] Diseases from the GWAS catalog were grouped according to the disease classification proposed by Wang *et al*.[41] Note that the "Digestive system" group includes Crohn's Disease, a subtype of Inflammatory Bowel Disease. [c] The enrichment *P*-value is the binomial tail probability for observing the shown number of TAGC SNPs with $P_{random} \leq 10^{-4}$ among the SNPs reported in the GWAS catalog for a group of diseases (for example, the probability of observing 20 or more SNPs with $P_{random} \leq 10^{-4}$ among the 743 cardiovascular SNP is shown in the last column). We investigated 15 groups and a conservative Bonferroni adjusted significance threshold for enrichment is 0.05/15=0.003.

Table 4. Enrichment of asthma risk SNPs[a] in promoter and enhancer marks and DNase I hypersensitivity sites

Type of regulatory elements[b]	Proportion of all cell types (blood cell types) showing enrichment with a given false discover rate (FDR)[c]	
	FDR ≤ 10%	FDR ≤ 5%
All promoter states	6% (26%)	0
Active promoter states	13% (33%)	0
All enhancer states	57% (100%)	43% (89%)
Active enhancer states	66% (100%)	50% (100%)
DNase I hypersensitivity sites	16% (50%)	12% (40%)

[a]Enrichment was assessed at 16 asthma-loci identified by this study (Table 1); the 6p21.3 and 6p21.32 loci that encompass HLA were excluded because of the high amount of variability and LD in this region. [b] Enhancer and promoter states were derived from integrated functional genomics data of 127 ROADMAP and ENCODE reference epigenomes in various cell types (including 27 leukocytes)[27]. DNase I hypersensitivity sites were identified in 51 cell types (including 10 leukocytes)[27]. [c] Empirical-*P*-values for enrichment were obtained using 10,000 Monte-Carlo simulations of random sets of SNPs matching the original set of asthma-associated SNPs[43]; Benjamini-Hochberg's FDR was calculated to correct for multiple testing (see Online Methods for more detail).

Legends

Figure 1. Results of the European-ancestry meta-analysis across 127,669 individuals.

(a–f) Manhattan plot **(a)** and regional plots for loci reaching genome-wide significance in the European-ancestry meta-analysis, including the 6p22.1 locus **(b)**, the 6p21.33 locus **(c)**, the 6q15 locus **(d)**, the 16p13.13 locus **(e)** and the 17q21.33 locus **(f)**. The x axis presents physical distance in megabases (build 37.3) and the y axis presents $-\log_{10}(P_{random})$ values for association statistics. The lead SNP (rsID) in the region is shown as a purple diamond; for remaining SNPs, the color indicates r^2 with the lead SNP. In **a,** $-\log_{10}(P_{random})$ values are for all SNPs passing quality control in at least two-thirds of the studies; the dashed horizontal line represents the genome-wide significance level ($P_{random}=10^{-8}$); the novel signals detected in the European-ancestry meta-analysis are shown in red, the known signals are in black and those that will reach significance in the multi-ancestry meta-analysis are in blue.

Figure 2. Results of the multi-ancestry meta-analysis across all 142,486 individuals.

(a–e) Manhattan plot (a) and regional plots for loci reaching genome-wide significance in the multi-ancestry meta-analysis, including the 5q31.3 locus (b), the 8q21.13 locus (c), the 10p14 locus (d) and the 12q13.3 locus (e). The x axis presents physical distance in megabases (build 37.3) and the y axis presents $-\log_{10}(P_{random})$ values for association statistics. The lead SNP (rsID) in the region is shown as a purple diamond; for remaining SNPs, the color indicates r^2 with the lead SNP. In a, $-\log_{10}(P_{random})$ values are for all SNPs passing quality control in at least two-thirds of the studies; the dashed horizontal line represents the genome-wide significance level ($P_{random}=10^{-8}$); the novel signals detected in the multi-ancestry meta-analysis are shown in red, the known signals are in black and those that were detected in the European-ancestry meta-analysis are in blue.

Figure 3. Relationships between asthma-associated loci using GRAIL.

Graphical representation of the connections between SNPs and corresponding genes with GRAIL $P<0.05$ (the 10p14 locus was ignored since the lead SNP could not be assigned to any gene in the region by GRAIL). Thicker and redder lines imply stronger literature-based connectivity.

Figure 1

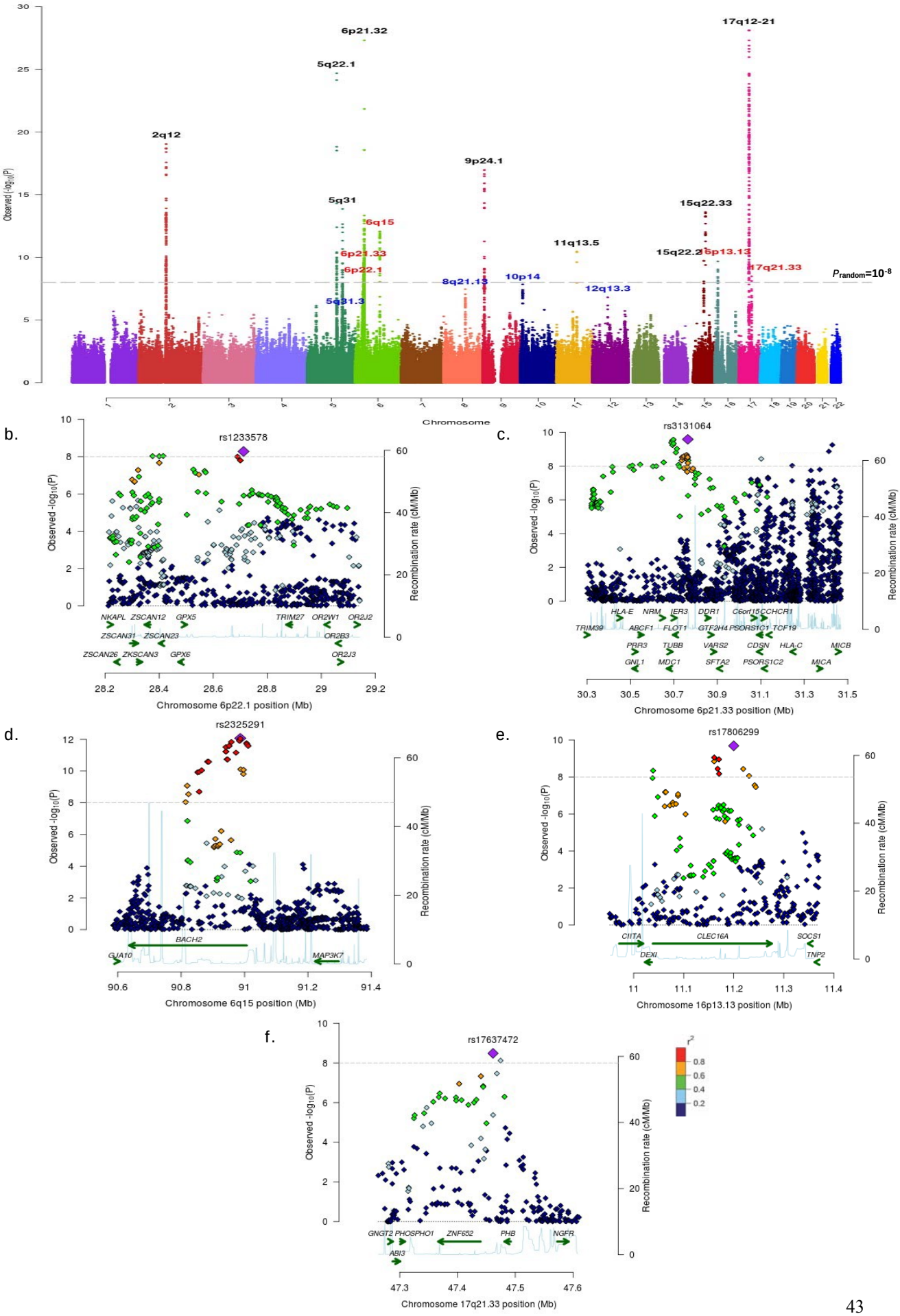

2q12
6p21.32
5q22.1
5q31
6q15
6p21.33
6p22.1
5q31.3
8q21.13
9p24.1
10p14
11q13.5
12q13.3
15q22.33
15q22.2
16p13.13
17q12-21
17q21.33
$P_{random}=10^{-8}$
Observed -log10(P)
Chromosome
b.
rs1233578
Chromosome 6p22.1 position (Mb)
c.
rs3131064
Chromosome 6p21.33 position (Mb)
d.
rs2325291
Chromosome 6q15 position (Mb)
e.
rs17806299
Chromosome 16p13.13 position (Mb)
f.
rs17637472
Chromosome 17q21.33 position (Mb)
Recombination rate (cM/Mb)

Figure 2

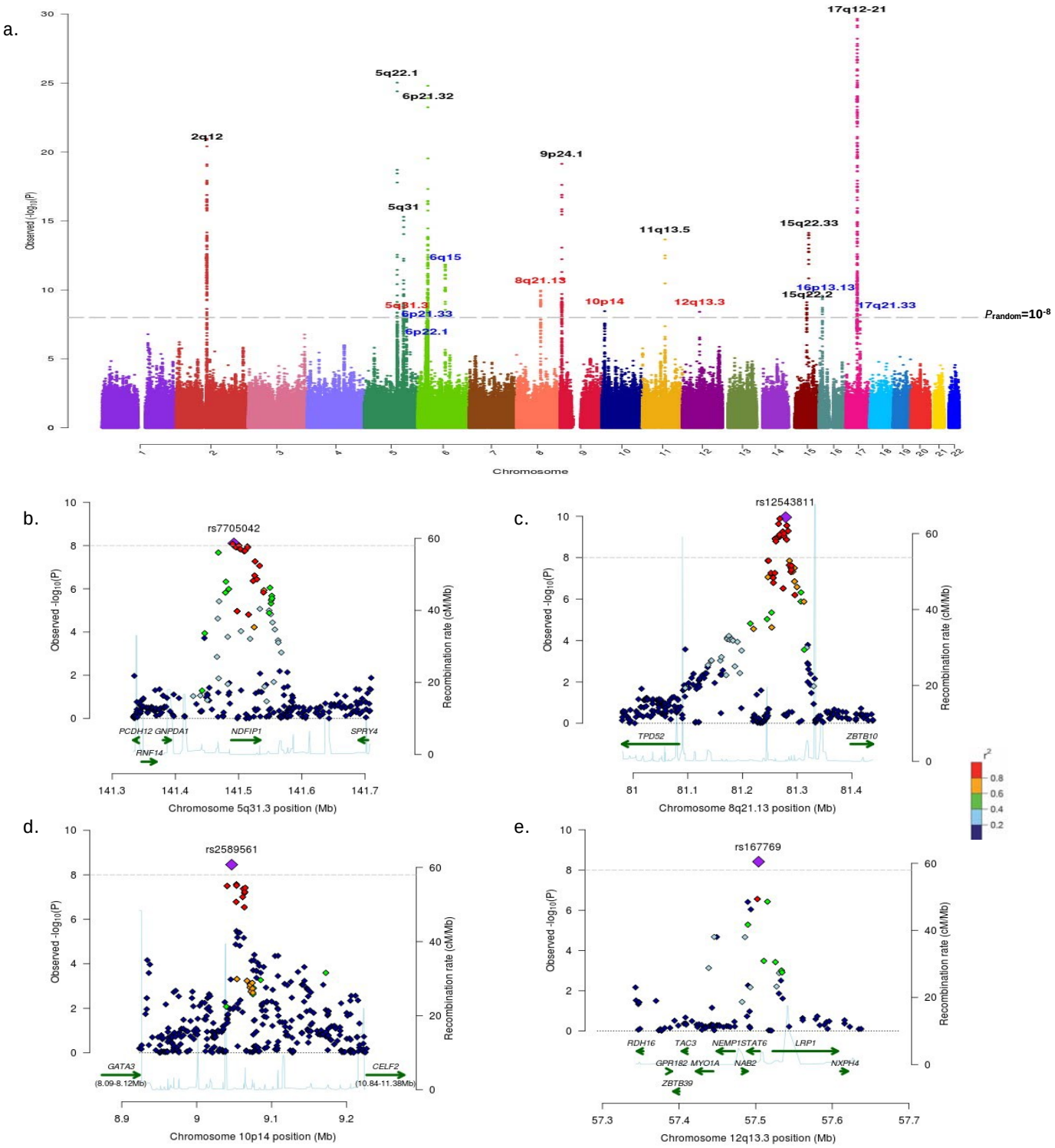
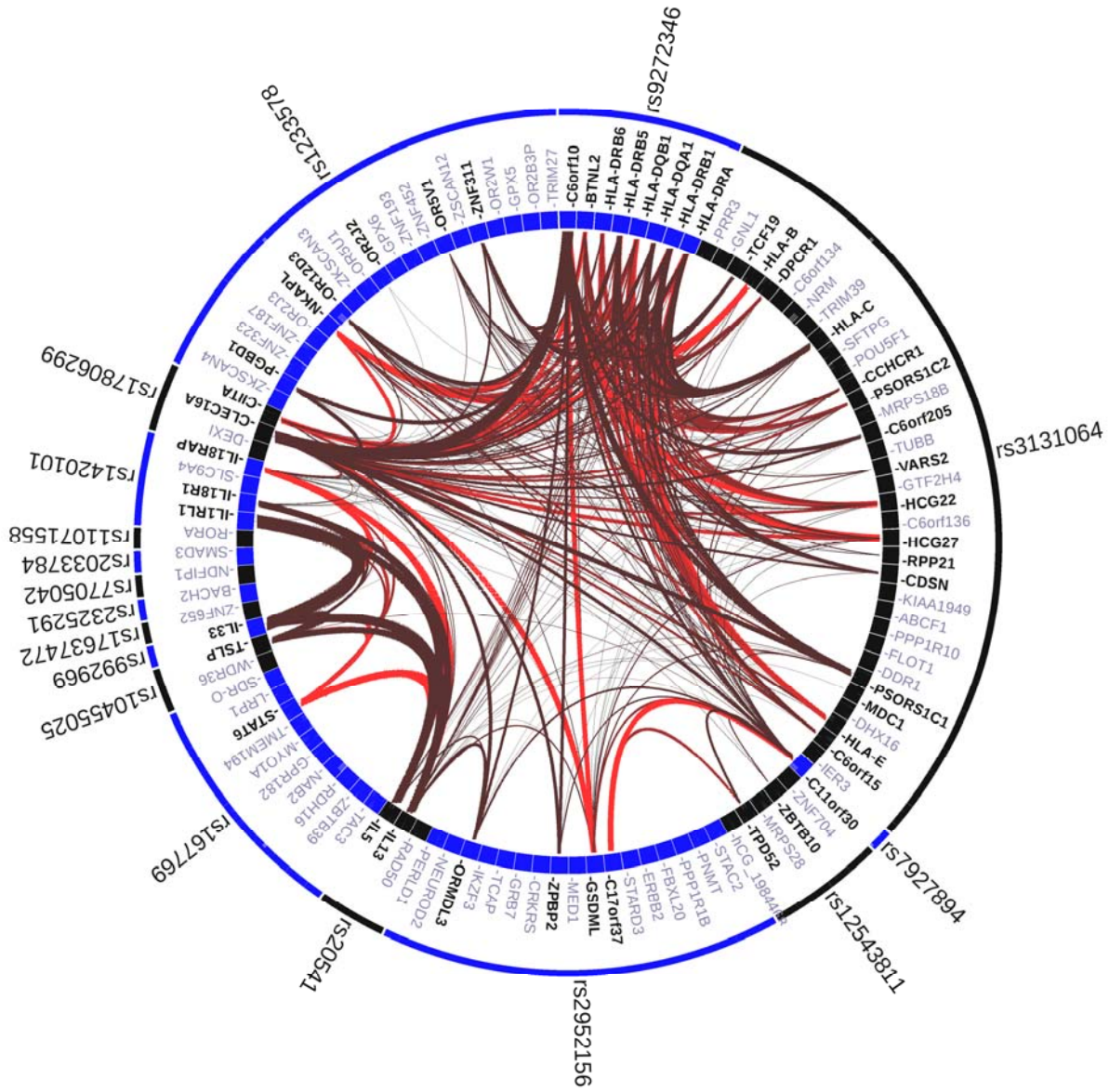

a.
Observed (-log10(P))
Chromosome
2q12
5q22.1
6p21.32
5q31
6q15
5q31.3
6p21.33
6p22.1
9p24.1
8q21.13
10p14
11q13.5
12q13.3
15q22.33
16p13.13
15q22.2
17q12-21
17q21.33
Prandom=10-8
b.
rs7705042
PCDH12 GNPDA1
RNF14
NDFIP1
SPRY4
Chromosome 5q31.3 position (Mb)
c.
rs12543811
TPD52
ZBTB10
Chromosome 8q21.13 position (Mb)
d.
rs2589561
GATA3
(8.09-8.12Mb)
CELF2
(10.84-11.38Mb)
Chromosome 10p14 position (Mb)
e.
rs167769
RDH16
TAC3
NEMP1
STAT6
LRP1
GPR182
MYO1A
NAB2
NXPH4
ZBTB39
Chromosome 12q13.3 position (Mb)
Observed -log10(P)
Recombination rate (cM/Mb)
r2

Figure 3


rs9272346
rs3131064
rs7927894
rs12543811
rs2952156
rs2054l
rs1677769
rs10455025
rs9929969
rs17637472
rs2325291
rs7705042
rs2033784
rs11071558
rs1420101
rs17806299
rs12335578
C6orf10
BTNL2
HLA-DRB6
HLA-DRB5
HLA-DQB1
HLA-DQA1
HLA-DRB1
HLA-DRA
PRR3
GNL1
TCF19
HLA-B
DPCR1
C6orf134
NRM
TRIM39
HLA-C
SFTPG
POU5F1
CCHCR1
PSORS1C2
MRPS18B
C6orf205
TUBB
VARS2
GTF2H4
HCG22
C6orf136
HCG27
RPP21
CDSN
KIAA1949
ABCF1
PPP1R10
FLOT1
DDR1
PSORS1C1
MDC1
DHX16
HLA-E
C6orf15
IER3
C11orf30
ZNF704
ZBTB10
MRPS28
TPD52
hCG_1984441
STAC2
PNMT
PPP1R1B
FBXL20
ERBB2
STARD3
C17orf37
GSDML
MED1
IKZF3
TCAP
GRB7
CRKRS
ZPBP2
ORMDL3
NEUROD2
PERLD1
RAD50
IL13
TAC3
ZBTB39
RPH3A
NAB2
GPR182
MYO1A
TMEM194
STAT6
LRP1
SDR-O
WDR36
TSLP
IL33
ZNF652
BACH2
NDFIP1
SMAD3
RORA
IL1RL1
IL18R1
SLC9A4
IL18RAP
DEXI
CLEC16A
CIITA
ZKSCAN4
PGBD1
ZNF323
ZNF187
OR2J3
NKAPL
OR12D3
ZKSCAN3
OR6U1
OR2I2
GPX6
ZNF192
ZNF452
ZSCAN22
OR5V1
ZNF311
OR2W1
GPX5
OR2B3P
TRIM27

References

1. Akinbami, L.J. *et al.* Trends in asthma prevalence, health care use, and mortality in the United States, 2001-2010. *NCHS Data Brief*, 1-8 (2012).
2. Duffy, D.L., Martin, N.G., Battistutta, D., Hopper, J.L. & Mathews, J.D. Genetics of asthma and hay fever in Australian twins. *Am Rev Respir Dis* **142**, 1351-8 (1990).
3. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**, D1001-6 (2014).
4. Torgerson, D.G. *et al.* Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet* **43**, 887-92 (2011).
5. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41-7 (2016).
6. Wain, L.V. *et al.* Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet* **49**, 416-425 (2017).
7. Bouzigon, E. *et al.* Effect of 17q21 variants and smoking exposure in early-onset asthma. *N Engl J Med* **359**, 1985-94 (2008).
8. Galanter, J.M. *et al.* Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: the Genes-environments & Admixture in Latino Americans study. *J Allergy Clin Immunol* **134**, 295-305 (2014).
9. Hirota, T. *et al.* Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat Genet* **43**, 893-6 (2011).
10. Ferreira, M.A. *et al.* Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. *J Allergy Clin Immunol* **133**, 1564-71 (2014).
11. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat Med* **21**, 1539-58 (2002).
12. Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* **44**, 1084-9 (2012).
13. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-43 (2013).
14. The GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
15. Liang, L. *et al.* A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. *Genome Res* **23**, 716-26 (2013).
16. Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* **8**, e1003029 (2012).
17. Nicodemus-Johnson, J. *et al.* DNA methylation in lung cells is associated with asthma endotypes and genetic risk. *JCI Insight* **1**, e90151 (2016).
18. Ding, L. *et al.* Rank-based genome-wide analysis reveals the association of ryanodine receptor-2 gene variants with childhood asthma among human populations. *Hum Genomics* **7**, 16 (2013).
19. Noguchi, E. *et al.* Genome-wide association study identifies HLA-DP as a susceptibility gene for pediatric asthma in Asian populations. *PLoS Genet* **7**, e1002170 (2011).
20. Bonnelykke, K. *et al.* A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat Genet* **46**, 51-5 (2014).
21. Himes, B.E. *et al.* Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. *Am J Hum Genet* **84**, 581-93 (2009).
22. Sleiman, P.M. *et al.* Variants of DENND1B associated with asthma in children. *N Engl J Med* **362**, 36-44 (2010).
23. Moffatt, M.F. *et al.* A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* **363**, 1211-21 (2010).
24. Ferreira, M.A. *et al.* Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet* **378**, 1006-14 (2011).

25. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75 (2012).
26. Zeller, T. *et al.* Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* **5**, e10693 (2010).
27. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
28. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet* **5**, e1000534 (2009).
29. Hong, S.W., Kim, S. & Lee, D.K. The role of Bach2 in nucleic acid-triggered antiviral innate immune responses. *Biochem Biophys Res Commun* **365**, 426-32 (2008).
30. Yang, M., He, R.L., Benovic, J.L. & Ye, R.D. beta-Arrestin1 interacts with the G-protein subunits beta1gamma2 and promotes beta1gamma2-dependent Akt signalling for NF-kappaB activation. *Biochem J* **417**, 287-96 (2009).
31. Cai, X. *et al.* Tripartite motif containing protein 27 negatively regulates CD4 T cells by ubiquitinating and inhibiting the class II PI3K-C2beta. *Proc Natl Acad Sci U S A* **108**, 20072-7 (2011).
32. Zurek, B. *et al.* TRIM27 negatively regulates NOD2 by ubiquitination and proteasomal degradation. *PLoS One* **7**, e41255 (2012).
33. Goenka, S. & Kaplan, M.H. Transcriptional regulation by STAT6. *Immunol Res* **50**, 87-96 (2011).
34. Qian, X., Gao, Y., Ye, X. & Lu, M. Association of STAT6 variants with asthma risk: a systematic review and meta-analysis. *Hum Immunol* **75**, 847-53 (2014).
35. Wang, Y., Tong, X. & Ye, X. Ndfip1 negatively regulates RIG-I-dependent immune signaling by enhancing E3 ligase Smurf1-mediated MAVS degradation. *J Immunol* **189**, 5304-13 (2012).
36. Venuprasad, K., Zeng, M., Baughan, S.L. & Massoumi, R. Multifaceted role of the ubiquitin ligase Itch in immune regulation. *Immunol Cell Biol* **93**, 452-60 (2015).
37. Javierre, B.M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369-1384 (2016).
38. Barnes, P.J. Pathophysiology of allergic inflammation. *Immunol Rev* **242**, 31-50 (2011).
39. Vicente, C.T. *et al.* Long-Range Modulation of PAG1 Expression by 8q21 Allergy Risk Variants. *Am J Hum Genet* **97**, 329-36 (2015).
40. Davison, L.J. *et al.* Long-range DNA looping and gene expression analyses identify DEXI as an autoimmune disease candidate gene. *Hum Mol Genet* **21**, 322-33 (2012).
41. Wang, L. *et al.* CPAG: software for leveraging pleiotropy in GWAS to reveal similarity between human traits links plasma fatty acids and intestinal inflammation. *Genome Biol* **16**, 190 (2015).
42. Rottem, M. & Shoenfeld, Y. Asthma as a paradigm for autoimmune disease. *Int Arch Allergy Immunol* **132**, 210-4 (2003).
43. Hayes, J.E. *et al.* Tissue-Specific Enrichment of Lymphoma Risk Loci in Regulatory Elements. *PLoS One* **10**, e0139360 (2015).
44. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).
45. Liang, L. *et al.* An epigenome-wide association study of total serum immunoglobulin E concentration. *Nature* **520**, 670-4 (2015).
46. Wenzel, S.E. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat Med* **18**, 716-25 (2012).
47. Smit, L.A. *et al.* 17q21 variants modify the association between early respiratory infections and asthma. *Eur Respir J* **36**, 57-64 (2010).
48. Caliskan, M. *et al.* Rhinovirus wheezing illness and genetic risk of childhood-onset asthma. *N Engl J Med* **368**, 1398-407 (2013).
49. Loss, G.J. *et al.* The Early Development of Wheeze. Environmental Determinants and Genetic Susceptibility at 17q21. *Am J Respir Crit Care Med* **193**, 889-97 (2016).
50. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control Clin Trials* **7**, 177-88 (1986).

51. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu Rev Genomics Hum Genet* **10**, 387-406 (2009).
52. Raychaudhuri, S. VIZ-GRAIL: visualizing functional connections across disease loci. *Bioinformatics* **27**, 1589-90 (2011).
53. So, H.C., Gui, A.H., Cherny, S.S. & Sham, P.C. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet Epidemiol* **35**, 310-7 (2011).